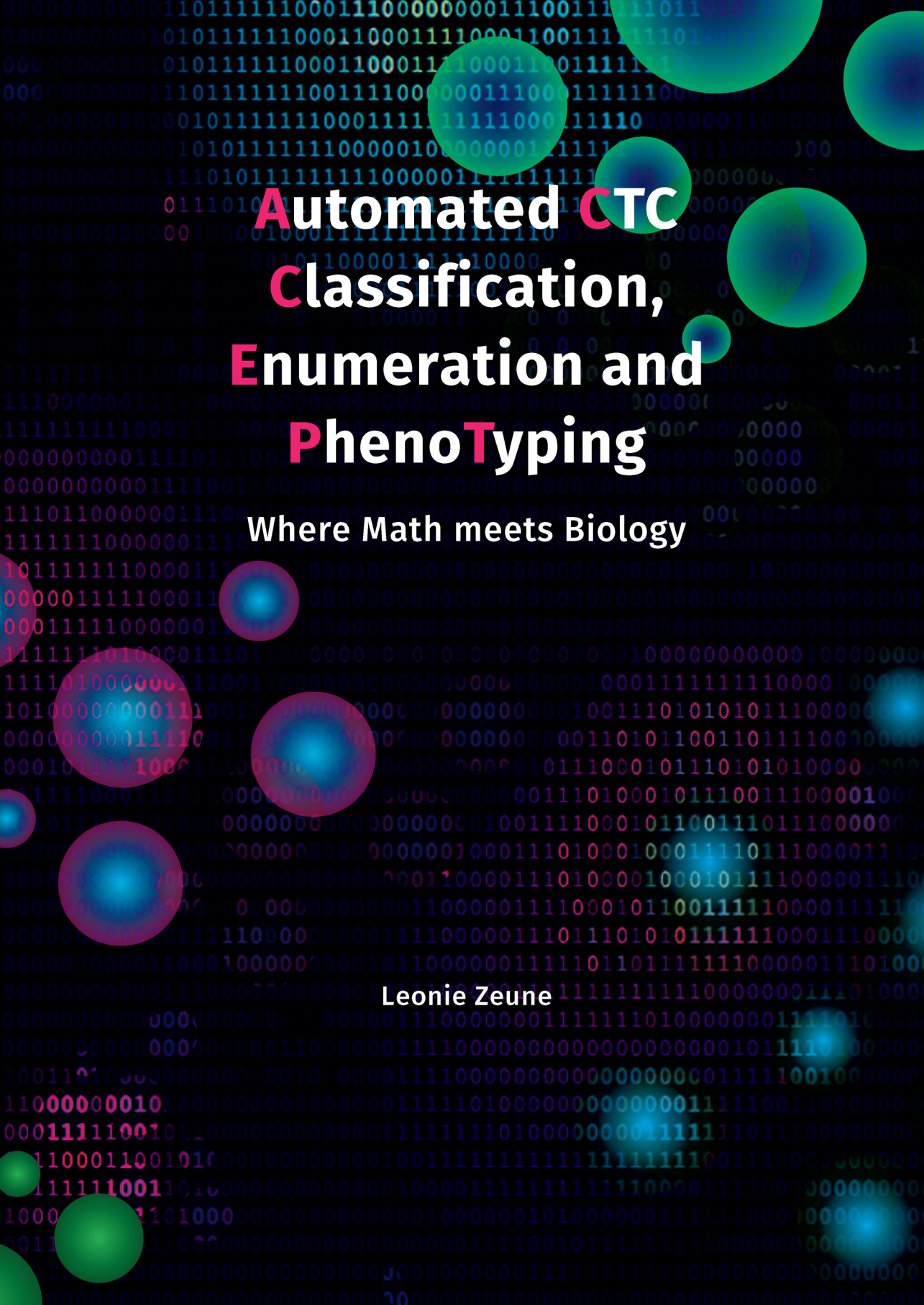

Automated CTC
Classification,
Enumeration and
PhenoTyping
Where Math meets Biology
Leonie Zeune

Automated CTC Classification, Enumeration and PhenoTyping

Where Math meets Biology

Leonie Laura Zeune

Automated CTC Classification, Enumeration and PhenoTyping

Dissertation

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. T.T.M. Palstra,
on account of the decision of the Doctorate Board,
to be publicly defended
on Thursday the 14th of February 2019 at 14:45 hours

by

Leonie Laura Zeune

born on 29th of May 1990
in Wuppertal, Germany

This dissertation has been approved by:

Supervisors:
prof. dr. L.W.M.M. Terstappen MD
prof. dr. S.A. van Gils

Co-supervisor:
dr. C. Brune

Members of the committee:

Chairman/Secretary:	
prof. dr. J.L. Herek	University of Twente
Supervisor:	
prof. dr. L.W.M.M. Terstappen, MD	University of Twente
Supervisor:	
prof. dr. S.A. van Gils	University of Twente
Co-supervisor:	
dr. C. Brune	University of Twente
Members:	
prof. dr. O. Öktem	KTH Stockholm
prof. dr. C.-B. Schönlieb	University of Cambridge
prof. dr. K. Pantel, MD	University Hospital Hamburg-Eppendorf
prof. dr. N. Stoecklein, MD	University Hospital Duesseldorf
prof. dr.ir. R.N.J. Veldhuis	University of Twente
prof. dr.ir. M.J.A.M. van Putten	University of Twente

UNIVERSITY OF TWENTE.

CANCER ID

CTCTrap

This work was financially supported by the EU FP-7 grant #305341 CTCTrap and the EU IMI project #115749-1 CANCER-ID.

ISBN 978-90-365-4713-0
DOI 10.3990/1.9789036547130

Contents

The subject of mathematics is so serious, that nobody should miss an opportunity to make it a little bit more entertaining.

Blaise Pascal

1 Introduction

Cancer is becoming the leading cause of death globally. In 2015, estimated 8.8 million people worldwide died from cancer and more than 14 million new cases occurred [53]. Thus, nearly one in six deaths was caused by cancer and for each of us the risk of developing cancer during our lifetime is around 38% thereby affecting most of our lives.

One of the major risk factors for developing cancer is age. As a further increase of life expectancy is likely, an even higher number of cancer patients is expected in the future. Early diagnosis and effective treatment methods are therefore of utmost importance to reduce mortality rates caused by cancer. This is also supported by the fact that, although the highest cancer incidence rates occur in high-income countries, around 70% of cancer deaths worldwide occur in low or middle-income countries providing less or worse medical care.

For most types of cancer, a biopsy, i.e. removal and microscopic analysis of cancer tissue, is the main instrument for diagnosis. A biopsy at the time of diagnosis is invasive and frequently non-repeatable through which no insight into disease progression and treatment effectiveness can be obtained. Another problem is that most cancer-related deaths are not caused by the primary tumor itself but by metastases at secondary sites. Although metastases share most of the characteristics of the primary tumor, they do have mutations that are not captured by biopsies from the primary tumor but might be important for treatment decisions.

Circulating Tumor Cells

The crucial point in the formation of metastases are cells that dissociate from the primary tumor and invade the bloodstream. These cells are called *Circulating Tumor Cells (CTCs)*. After entering the bloodstream, CTCs circulate through the cardiovascular system. The majority will be destroyed but some CTCs can extravasate the bloodstream at a secondary site and form new tumors called metastases. Figure 1 shows a schematic overview of this process.

CTCs can be found in the blood of all metastatic cancer patients provided a large enough blood volume as samples [11] and are predicted to be present before distant metastases have formed [12]. The genetic and phenotypic make-up of a CTC shows a large heterogeneity

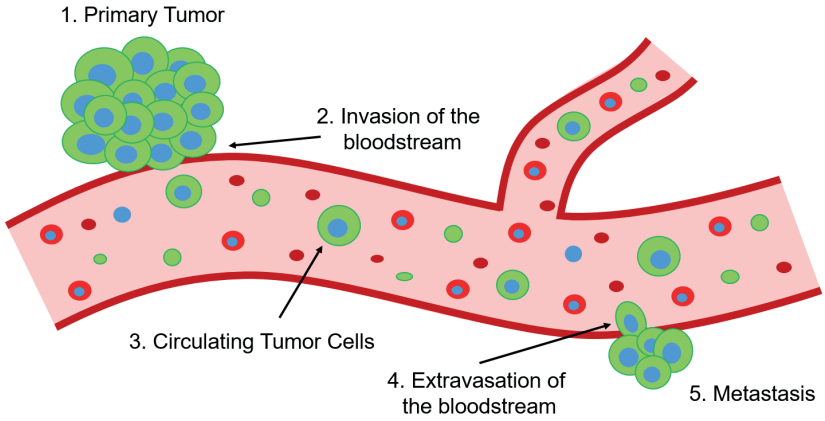


Figure 1: *Schematic overview of CTCs entering the bloodstream.*

[10, 34, 35, 54]. Notably, CTCs reflect not only characteristics of the primary biopsy but also changes seen in the metastatic sites and can thus be considered as a "liquid biopsy" covering all tumor sites. Opposite to tissue biopsies, obtaining CTCs from a patient is non-invasive and these liquid biopsies are repeatable. Thus, CTCs can be used to monitor the disease status of a patient and whether a therapy is working or not. Potentially CTCs can even be used to guide treatment decisions, thereby paving the way for personalized medicine for cancer patients.

Although CTCs offer great potential as a biomarker for cancer, using this potential is partly hindered by their challenging detection. In comparison to other blood cells, CTCs are very rare and in 50% of metastatic patients only 1-10 CTCs can be found in 1 mL of blood. On the other hand, 1 mL of blood contains around $5 \cdot 10^6$ white blood cells (WBCs) and even around $5 \cdot 10^9$ red blood cells, therefore making it very challenging to detect CTCs among all cells.

To detect and count CTCs in the blood of patients, they need to be separated from the bulk of other cells in the blood. Currently, the only system that is cleared by the FDA for monitoring patients is the CellSearch® system (Menarini Silicon Biosystems Inc., Huntingdon Valley, PA, USA) which was shown to identify CTCs in patients with lung, prostate, colon, breast and pancreatic cancer [2]. Here, ferrofluids (175 nm magnetic nanoparticles) coated with EpCAM antibodies are used to separate CTCs from most other cells using magnets. Afterwards the CTCs and remaining leukocytes (roughly $1 - 100 \cdot 10^3$ cells) are fluorescently stained, highlighting the cell nucleus and the cytoskeleton. An automated fluorescent scanning microscope is used to image all cells, generating around 180 images per fluorophore per sample. Based on their fluorescent expression, CTC candidates (objects that stain with the nucleic acid dye and cytokeratin) are presented to the user in a thumbnail gallery and are manually reviewed, making the CTC counts subjective. The same holds true for most other techniques – although more and more technologies for CTC isolation are recently developed [1, 3, 5, 46], most of them lack a (semi-) automated image processing pipeline to generate CTC counts.

The importance of accurate CTC counts Notably, the number of CTCs enumerated in 7.5 mL of blood is directly related to the survival prospects of metastatic cancer patients. In primary cancer, their presence is associated with an increased risk for disease recurrence and survival. This was demonstrated in prospective multi-center studies for metastatic breast, colorectal and prostate cancer [9, 13, 15] and clinical utility was shown for even more cancer types [18, 24, 25, 27, 31, 45, 47, 56]. Moreover, a variety of studies showed that treatment targets on CTCs can be detected and used to guide patient's therapy [4, 14, 23, 41, 42, 43, 52]. In the original studies the discrimination of patients with favorable versus unfavorable prognosis was based on a fixed cut-off value, namely three CTCs for colorectal or five CTCs for breast and prostate cancer. These small cut-offs show that it is of utmost importance to accurately assign objects as CTC, especially for monitoring therapy based on CTC counts and to diagnose the presence of cancer beyond the primary tumor.

Nevertheless, the reported CTC counts are still subjective and prone to inter-reader differences. It was shown that extensive training can reduce the variations in assigning objects as CTCs, but it is not possible to completely eliminate them as long as the process is not fully automated and objective [26, 30, 55]. This strongly motivates the development of *fully automated computer algorithms* to *detect and count CTCs* in the blood probes of metastatic patients.

Key facts:

- Cancer is becoming the leading cause of death globally
- Most patients die from metastasis formed by *Circulating Tumor Cells*
- The number of CTCs in patients' blood is directly related to their survival chances
- CTCs can be used to personalize treatments based on their number and characteristics
- Currently there are no fully automated and unbiased algorithms to analyze CTCs

Towards Personalized Cancer Medicine

Personalized medicine also known as "theranostics" [7, 22, 29] is nowadays one of the main objectives in our health care system. The goal is to move away from a "one therapy fits all" approach for patients with the same diagnosis and move to individualized treatment and medication based on the needs of the individual patient. For the patient, this would result in improved treatment efficacy, higher recovery chances and less burden from inefficient treatments. But also from an economical view point, personalized medicine is reasonable in order to lower the costs of treatments that are not beneficial for the patient. The idea of personalized medicine is based on the fact that every person's human genome is unique and influences the way patients react to treatment options.

In the field of cancer research, liquid biopsies like CTCs and cell-free tumor DNA (ctDNA) are a step towards personalized treatments. They can be used to obtain a comprehensive picture of a patient and gain more insight into the patient's cancer characteristics. The subsequent treatment can then be tailored to the patient's needs. The formation of cancer is based on a random accumulation of DNA mutations and epigenetic alterations - finding these alterations is therefore needed to choose a treatment specifically targeting them.

One key ingredient for personalized medicine is the availability of large data sets that can be used to base a decision on. Over the years, the amount of data that is recorded and stored in the biomedical field is constantly growing. Moreover, research institutes worldwide become more and more globally connected and data is shared worldwide, including medical databases like "The Cancer Genome Atlas" (https://cancergenome.nih.gov/) for genetic mutations or the "National Biomedical Imaging Archive" (http://imaging.nci.nih.gov) for medical images. Yet, the availability of large datasets inevitably leads to the challenge of analyzing the data which is often impossible without the use of automated analysis tools. This is the point where artificial intelligence (AI) techniques can help. Machine learning (ML) techniques, a subclass of AI based on statistical methods, have the power to analyze large datasets efficiently and fully automated. Therefore, they could provide the doctors with only the most important information to base their decision on. Moreover, techniques like sparse coding are able to find hidden patterns in the data that were not directly evident but might provide useful information. Thus, to really use the collected data, it needs machine learning combined with expert knowledge on what the optimal model and strategy is to analyze given data [44].

The ongoing success of machine learning techniques in the biomedical field led to numerous publications in the previous decades. In various biomedical applications, AI could not only compete with human operators, but overcame the threshold of human analytical performance, see e.g. [17, 20, 32, 39]. Automated data-driven diagnosis and disease prediction can pave the way for personalized medicine and thereby lift healthcare to a new level of quality and understanding. Especially the development of machine learning methods based on convolutional neural networks (CNNs) and deep learning (DL) [36] revolutionizes the biomedical imaging field, reaching from organ level imaging like radiology [19, 58] to single cell imaging. See [37, 8] for two recent reviews about DL in biomedicine. Two fields where ML methods are quickly gaining importance are radiomics and histopathology. Radiomics describes the automated extraction and analysis of a large amount of quantitative image features from radiology images, originating for example from CT, PET or MRI machines [33]. The goal of radiomics is to unravel the image features that are most important for diagnosis and prognosis of disease progression by automated learning from large amounts of data. In histopathology, CNNs have strongly improved the automated discrimination of cancer tissue from healthy tissue [49, 50, 28, 57, 38]. Due to a relatively easy access to large amounts of training data via biopsies, this field has strongly profited from deep learning techniques. Reducing the imaging size scale even further, ML methods are also applied for single cell detection [60, 63, 61] and single cell classification or profiling methods [48, 59, 16, 6] via deep learning. Although AI in biomedicine resulted already in many breakthrough discoveries, it was shown that often the best performance is achieved by combining the power of

machine learning with the expert knowledge of human operators [58]. Yet, in many cases, missing collaboration between doctors on the one hand and data scientists on the other hand prevents that the power of big data analysis is fully exploited for the patient's benefit.

The combination of machine learning and expert knowledge also allows for a very high reliability. For human operators the main sources of error are negligence, fatigue and diminishing concentration where a computer is resistant to. On the other hand, for machine learning algorithms it is often not fully understood how the algorithm made the decision, some methods are seen as black boxes that need to be revealed. Common strategies to understand their working principle are either to build models that can be related to existing mathematical theory [21, 40] or to visualize the way a machine learning algorithm "looks" at the data [51, 62]. The goal is to minimize so called "adversarial cases" where a small change in the input data leads to a wrong output which is not directly explainable from the human point of view. Finding and understanding these cases, even in the case of high overall accuracy, is therefore very important to improve the algorithm's reliability and expert knowledge is needed to guide machine learning algorithms in the correct direction.

From a mathematical point of view, the most important points to address in order to provide reliable data analysis are therefore robustness, reproducability and uncertainty quantification. This should prevent that small alterations in the input data lead to large deviations in the output. Moreover, it is important to quantify the uncertainty to get insight into the reliability of algorithms and to use this to compare different methods not only based on performance. In this thesis we will therefore focus on answering the following questions:

- How can we construct robust segmentation methods that can reliably find objects with varying intensity and shape?
- What is the influence of different visualizations on manual cell classifications and how can quantifiable features minimize the variation?
- Which excess profit can we gain from automated cell analysis and can we gain new biomedical insight in the field of CTCs that was hidden before?
- How can we automate the detection and classification of CTCs and can we get a performance comparable to human performance?
- Where are the hidden uncertainties in Deep Learning approaches for CTC classification and which tools can we use to quantify and minimize them?

Answers to these questions are provided in this thesis and are incorporated in our open-source toolbox for CTC research, called "ACCEPT – Automated CTC Classification, Enumeration and PhenoTyping" which can be downloaded from *https://github.com/LeonieZ/ACCEPT*. With this toolbox we aim to provide an advanced but user-friendly cell analysis tool to the research field. The toolbox can be downloaded for free and provides tools to 1. analyze CTCs scored beforehand, 2. detect and analyze all cell events in fluorescent images of blood probes, 3. provide an automated classification into CTCs and other cell classes and 4. present all results in a thumbnail gallery combined with multiple scatter plots for evaluating cell features. With this toolbox we want to make a first step towards closing the gap between automated imaging and machine learning science and biomedical research in the field of liquid biopsies.

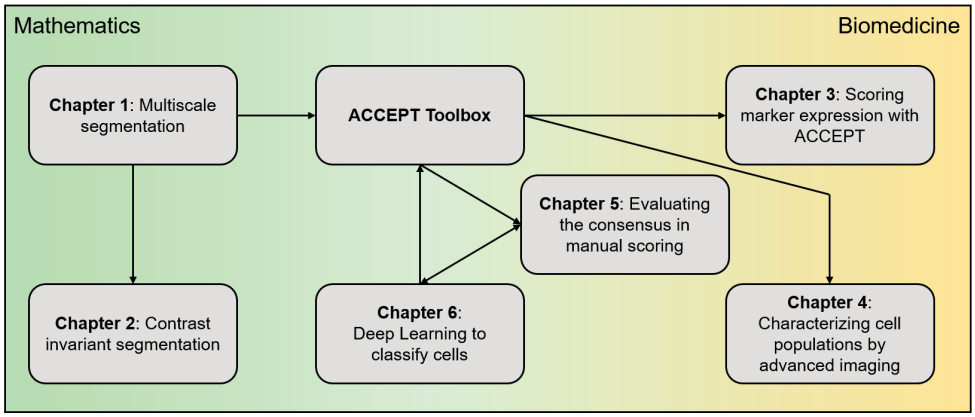


Figure 2: *Schematic overview of the thesis outline.*

Thesis Outline

This thesis starts with a brief introduction of the research concepts from the fields of mathematical imaging and machine learning that are used throughout the thesis. The following six chapters present the research we performed throughout the PhD project. Most concepts are either incorporated in our open-source toolbox "ACCEPT" or are based on the use of the toolbox. In Figure 2 we give a schematic overview of the chapters and how they are related to each other and to the ACCEPT toolbox. Here, the papers are sorted from more mathematical papers on the left hand side to biomedical papers on the right hand side. The project is centered around the development of ACCEPT - the theoretical papers are input to the toolbox and the applied papers are output of the toolbox.

In Chapter 1, we introduce a multiscale segmentation model. With this model, we can reliably detect objects of different shapes and intensities. Together with the segmentation, we obtain a so called "spectral response" function, which contains information about the frequencies of objects with varying size and intensity in an image. Moreover, we can show that this model is very robust to noise and applicable to a wide range of images. This model is used in our open-source toolbox ACCEPT to detect CTCs and other cell-types in fluorescent images of blood probes. In Chapter 2, we adapt the multiscale segmentation model presented before in order to distinguish different objects only based on their size and not based on their intensity. In that way, we eliminate the ambiguity between size and intensity changes in the spectral response function seen before. Chapter 3 demonstrates a first application of ACCEPT to detect treatment targets on prescored CTCs. Here, we analyze CTCs from metastatic breast cancer patients with regard to HER-2 expression. Over-expression of HER-2 is strongly associated with poor prognosis and increased disease recurrence and can be used as a biomarker to base the treatment on. We show that the ACCEPT toolbox can help to reduce the user-bias in deciding whether a CTC is HER-2 positive or not. In the subsequent Chapter 4, we present another application of the ACCEPT toolbox. We use the toolbox to detect all cells present in the fluorescent images of the blood of metastatic non-

small cell lung cancer patients and healthy controls. Based on the morphological features extracted, we group them into different cell classes. We show that not only CTCs are more frequent in the blood of cancer patients, but all nucleated events. The overall increase of cells leads to very clustered cell images and we evaluate whether more single cells can be detected with Deep Learning based segmentation methods compared to the multiscale approach from Chapter 1. Chapter 5 examines the user-bias in scoring cells as a CTC. For this purpose, we created a set of 100 cells and asked 15 reviewers whether the cells are a CTC or not. Moreover, we asked an expert panel to mutually score the same set of cells. As a third comparison, we used the results of a Deep Learning classification network for CTC Scoring. Our comparison shows that the agreement between the reviewers and the classification algorithm is equally high as the agreement with the expert panel. This motivates the work in Chapter 6 where we train a Deep Learning classification network to distinguish five different cell classes commonly found in the blood of metastatic cancer patients. We evaluate how more advanced network architectures can improve the reliability of the automated decision and how visualization can help to find new subclasses of cells. Afterwards, we conclude this thesis and provide an outlook to future research directions.

Bibliography

[1] C. Alix-Panabières and K. Pantel. Challenges in Circulating Tumour Cell Research. *Nature Reviews Genetics*, 14(9):623–631, 2014.

[2] W. J. Allard, J. Matera, M. C. Miller, M. I. Repollet, M. C. Connelly, C. G. Rao, A. G. Tibbe, J. W. Uhr, and L. W. Terstappen. Tumor Cells Circulate in the Peripheral Blood of All Major Carcinomas but not in Healthy Subjects or Patients with Nonmalignant Diseases. *Clinical Cancer Research*, 10(20):6897–6904, 2004.

[3] W. J. Allard and L. W. Terstappen. CCR 20th Anniversary Commentary: Paving the Way for Circulating Tumor Cells. *Clinical Cancer Research*, 21(13):2883–2885, 2015.

[4] G. Attard, J. F. Swennenhuis, D. Olmos, A. H. Reid, E. Vickers, R. A'Hern, R. Levink, F. A. Coumans, J. Moreira, R. Riisnaes, N. B. Oommen, G. Hawche, C. Jameson, E. Thompson, R. C. Sipkema, C. P. Carden, C. Parker, D. Dearnaley, S. B. Kaye, C. S. Cooper, A. Molina, M. E. Cox, L. W. Terstappen, and J. S. de Bono. Characterization of ERG, AR and PTEN Gene Status in Circulating Tumor Cells from Patients with Castration-Resistant Prostate Cancer. *Cancer Research*, 69(7):2912–2918, 2009.

[5] A. M. Barradas and L. W. Terstappen. Towards the Biological Understanding of CTC: Capture Technologies, Definitions and Potential to Create Metastasis. *Cancers*, 5(4):1619–1642, 2013.

[6] J. C. Caicedo, S. Cooper, F. Heigwer, S. Warchal, P. Qiu, C. Molnar, A. S. Vasilevich, J. D. Barry, H. S. Bansal, O. Kraus, M. Wawer, L. Paavolainen, M. D. Herrmann, M. Rohban, J. Hung, H. Hennig, J. Concannon, I. Smith, P. A. Clemons, S. Singh, P. Rees, P. Horvath, R. G. Linington, and A. E. Carpenter. Data-Analysis Strategies for Image-Based Cell Profiling. *Nature Methods*, 14(9):849–863, aug 2017.

[7] L. Chin, J. N. Andersen, and P. A. Futreal. Cancer Genomics: From Discovery Science to Personalized Medicine. *Nature Medicine*, 17(3):297, 2011.

[8] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. DeCaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter, and C. S. Greene. Opportunities and Obstacles for Deep Learning in Biology and Medicine. *Journal of The Royal Society Interface*, 15(141):20170387, apr 2018.

[9] S. J. Cohen, C. J. Punt, N. Iannotti, B. H. Saidman, K. D. Sabbath, N. Y. Gabrail, J. Picus, M. Morse, E. Mitchell, M. C. Miller, G. V. Doyle, H. Tissing, L. W. Terstappen, and N. J. Meropol. Relationship of Circulating Tumor Cells to Tumor Response, Progression-Free Survival, and Overall Survival in Patients with Metastatic Colorectal Cancer. *Journal of Clinical Oncology*, 26:3213–3221, 2008.

[10] F. A. Coumans, C. J. Doggen, G. Attard, J. S. de Bono, and L. W. Terstappen. All Circulating EpCAM+ CK+ CD45- Objects Predict Overall Survival in Castration-Resistant Prostate Cancer. *Annals of Oncology*, 21(9):1851 – 1857, 2010.

[11] F. A. Coumans, S. T. Ligthart, and L. W. Terstappen. Interpretation of Changes in Circulating Tumor Cell Counts. *Translational Oncology*, 5(6):486–491, 2012.

[12] F. A. Coumans, S. Siesling, and L. W. Terstappen. Detection of Cancer Before Distant Metastasis. *BMC Cancer*, 13(1):283, 2013.

[13] M. Cristofanilli, G. T. Budd, M. J. Ellis, A. Stopeck, J. Matera, M. C. Miller, J. M. Reuben, G. V. Doyle, W. J. Allard, L. W. Terstappen, and D. F. Hayes. Circulating Tumor Cells, Disease Progression, and Survival in Metastatic Breast Cancer. *New England Journal of Medicine*, 351(8):781–791, 2004.

[14] J. S. de Bono, G. Attard, A. Adjei, M. N. Pollak, P. C. Fong, P. Haluska, L. Roberts, C. Melvin, M. I. Repollet, D. A. Chianese, M. C. Connelly, L. W. Terstappen, and A. Gualberto. Potential Applications for Circulating Tumor Cells Expressing the Insulin-Like Growth Factor-I Receptor. *Clinical Cancer Research*, 13(12):3611–3616, 2007.

[15] J. S. de Bono, H. I. Scher, R. B. Montgomery, C. Parker, M. C. Miller, H. Tissing, G. V. Doyle, L. W. Terstappen, K. J. Pienta, and D. Raghavan. Circulating Tumor Cells Predict Survival Benefit From Treatment in Metastatic Castration-Resistant Prostate Cancer. *Clinical Cancer Research*, 14(19):6302–6309, 2008.

[16] O. Dürr and B. Sick. Single-Cell Phenotype Classification Using Deep Convolutional Neural Networks. *Journal of Biomolecular Screening*, 21(9):998–1003, oct 2016.

[17] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun. Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks. *Nature*, 542(7639):115–118, 2017.

[18] P. Gazzaniga, A. Gradilone, E. de Berardinis, G. M. Busetto, C. Raimondi, O. Gandini, C. Nicolazzo, A. Petracca, B. Vincenzi, A. Farcomeni, V. Gentile, E. Cortesi, and L. Frati. Prognostic Value of Circulating Tumor Cells in Nonmuscle Invasive Bladder Cancer: a CellSearch Analysis. *Annals of Oncology*, 23(9):2352–2356, 2012.

[19] H. Greenspan, B. van Ginneken, and R. M. Summers. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging*, 35(5):1153–1159, may 2016.

[20] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. *JAMA - Journal of the American Medical Association*, 316(22):2402–2410, 2016.

[21] E. Haber and L. Ruthotto. Stable Architectures for Deep Neural Networks. *Inverse Problems*, 34(1):014004, 2017.

[22] M. A. Hamburg and F. S. Collins. The Path to Personalized Medicine. *The New England Journal of Medicine*, 363(4):301–304, 2010.

[23] D. F. Hayes, T. M. Walker, B. Singh, E. S. Vitetta, J. W. Uhr, S. Gross, C. G. Rao, G. V. Doyle, and L. W. Terstappen. Monitoring Expression of HER-2 on Circulating Epithelial Cells in Patients with Advanced Breast Cancer. *International Journal of Oncology*, 21(5):1111–1117, 2002.

I

[24] T. J. N. Hiltermann, M. M. Pore, A. van den Berg, W. Timens, H. M. Boezen, J. J. Liesker, J. H. Schouwink, G. J. Wijnands, G. S. Kerner, F. A. Kruyt, H. Tissing, A. G. Tibbe, L. W. Terstappen, and H. J. Groen. Circulating Tumor Cells in Small-Cell Lung Cancer: a Predictive and Prognostic Factor. *Annals of Oncology*, 23(11):2937–2942, 2012.

[25] K. Hiraiwa, H. Takeuchi, H. Hasegawa, Y. Saikawa, K. Suda, T. Ando, K. Kumagai, T. Irino, T. Yoshikawa, S. Matsuda, M. Kitajima, and Y. Kitagawa. Clinical Significance of Circulating Tumor Cells in Blood from Patients with Gastrointestinal Cancers. *Annals of Surgical Oncology*, 15(11):3092, 2008.

[26] M. Ignatiadis, S. Riethdorf, F.-C. Bidard, I. Vaucher, M. Khazour, F. Rothé, J. Metallo, G. Rouas, R. E. Payne, R. Coombes, I. Teufel, U. Andergassen, S. Apostolaki, E. Politaki, D. Mavroudis, S. Bessi, M. Pestrin, A. Di Leo, M. Campion, M. Reinholz, E. Perez, M. Piccart, E. Borgen, B. Naume, J. Jimenez, C. Aura, L. Zorzino, M. Cassatella, M. Sandri, B. Mostert, S. Sleijfer, J. Kraan, W. J. Janni, T. Fehm, B. Rack, L. W. Terstappen, M. I. Repollet, J.-Y. Pierga, M. C. Miller, C. Sotiriou, S. Michiels, and K. Pantel. International Study on Inter-Reader Variability for Circulating Tumor Cells in Breast Cancer. *Breast Cancer Research*, 16(2):R43, 2014.

[27] W. J. Janni, B. Rack, L. W. Terstappen, J.-Y. Pierga, F.-A. Taran, T. Fehm, C. Hall, M. R. de Groot, F.-C. Bidard, T. W. Friedl, P. A. Fasching, S. Y. Brucker, K. Pantel, and A. Lucci. Pooled Analysis of the Prognostic Relevance of Circulating Tumor Cells in Primary Breast Cancer. *Clinical Cancer Research*, 22(10):2583–2593, 2016.

[28] A. Kapil, A. Meier, A. Zuraw, K. Steele, M. Rebelatto, G. Schmidt, and N. Brieu. Deep Semi Supervised Generative Learning for Automated PD-L1 Tumor Cell Scoring on NSCLC Tissue Needle Biopsies. *arXiv:1806.11036*, jun 2018.

[29] S. S. Kelkar and T. M. Reineke. Theranostics: Combining Imaging and Therapy. *Bioconjugate Chemistry*, 22(10):1879–1903, 2011.

[30] J. Kraan, S. Sleijfer, M. H. Strijbos, M. Ignatiadis, D. Peeters, J.-Y. Pierga, F. Farace, S. Riethdorf, T. Fehm, L. Zorzino, A. G. Tibbe, M. Maestro, R. Gisbert-Criado, G. Denton, J. S. de Bono, C. Dive, J. A. Foekens, and J. W. Gratama. External Quality Assurance of Circulating Tumor Cell Enumeration Using the CellSearch® System: A Feasibility Study. *Cytometry Part B*, 80B(2):112–118, 2011.

[31] M. G. Krebs, R. Sloane, L. Priest, L. Lancashire, J.-M. Hou, A. Greystoke, T. H. Ward, R. Ferraldeschi, A. Hughes, G. Clack, M. Ranson, C. Dive, and F. H. Blackhall. Evaluation and Prognostic Significance of Circulating Tumor Cells in Patients with Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology*, 29(12):1556–1563, 2011.

[32] P. Lakhani and B. Sundaram. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*, 284(2):574–582, 2017.

[33] P. Lambin, R. T. Leijenaar, T. M. Deist, J. Peerlings, E. E. de Jong, J. van Timmeren, S. Sanduleanu, R. T. Larue, A. J. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. M. Mottaghy, J. E. Wildberger, and S. Walsh.

Radiomics: the Bridge Between Medical Imaging and Personalized Medicine. *Nature Reviews Clinical Oncology*, 14(12):749, oct 2017.

[34] M. B. Lambros, G. Seed, S. Sumanasuriya, V. Gil, M. Crespo, M. Sousa Fontes, R. Chandler, N. Mehra, G. Fowler, B. Ebbs, P. Flohr, S. Miranda, W. Yuan, A. Mackay, A. Ferreira, R. Pereira, C. Bertan, I. Figueiredo, R. Riisnaes, D. Nava-Rodrigues, A. Sharp, J. Goodall, G. Boysen, S. Carreira, D. Bianchini, P. Rescigno, Z. Zafeiriou, J. Hunt, D. Moloney, L. Hamilton, R. P. Neves, J. F. Swennenhuis, K. C. Andree, N. H. Stoecklein, L. W. Terstappen, and J. S. de Bono. Single Cell Analyses of Prostate Cancer Liquid Biopsies Acquired by Apheresis. *Clinical Cancer Research*, jan 2018.

[35] C. J. Larson, J. G. Moreno, K. J. Pienta, S. Gross, M. I. Repollet, S. M. O'hara, T. Russell, and L. W. Terstappen. Apoptosis of Circulating Tumor Cells in Prostate Cancer Patients. *Cytometry Part A*, 62A(1):46–53, 2004.

[36] Y. A. LeCun, Y. Bengio, and G. E. Hinton. Deep Learning. *Nature*, 521(7553):436, 2015.

[37] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. Arindra Adiyoso Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sanchez. A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, 42:60–88, 2017.

[38] G. Litjens, C. I. Sánchez, N. Timofeeva, M. Hermsen, I. D. Nagtegaal, I. Kovacs, C. Hulsbergen - van de Kaa, P. Bult, B. van Ginneken, and J. A. van der Laak. Deep Learning as a Tool for Increased Accuracy and Efficiency of Histopathological Diagnosis. *Scientific Reports*, 6(1):26286, sep 2016.

[39] J. Liu, B. Xu, L. Shen, J. Garibaldi, and G. Qiu. HEp-2 Cell Classification Based on a Deep Autoencoding-Classification Convolutional Neural Network. In *IEEE International Symposium on Biomedical Imaging*, pages 1019–1023. IEEE, 2017.

[40] S. Mallat. Understanding Deep Convolutional Networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150203, 2016.

[41] M. Mazel, W. Jacot, K. Pantel, K. Bartkowiak, D. Topart, L. Cayrefourcq, D. Rossille, T. Maudelonde, T. Fest, and C. Alix-Panabières. Frequent Expression of PD-L1 on Circulating Breast Cancer Cells. *Molecular Oncology*, 9(9):1773–1782, 2015.

[42] S. Meng, D. Tripathy, S. Shete, R. Ashfaq, B. Haley, S. Perkins, P. Beitsch, A. Khan, D. Euhus, C. Osborne, E. Frenkel, S. Hoover, M. Leitch, E. Clifford, E. S. Vitetta, L. Morrison, D. Herlyn, L. W. Terstappen, T. Fleming, T. Fehm, T. Tucker, N. Lane, J. Wang, and J. W. Uhr. HER-2 Gene Amplification Can Be Acquired as Breast Cancer Progresses. *Proceedings of the National Academy of Sciences*, 101(25):9393–9398, 2004.

[43] S. Meng, D. Tripathy, S. Shete, R. Ashfaq, H. Saboorian, B. Haley, E. Frenkel, D. Euhus, M. Leitch, C. Osborne, E. Clifford, S. Perkins, P. Beitsch, A. Khan, L. Morrison, D. Herlyn, L. W. Terstappen, N. Lane, J. Wang, and J. W. Uhr. uPAR and HER-2 Gene Status in Individual Breast Cancer Cells from Blood and Tissues. *Proceedings of the National Academy of Sciences*, 103(46):17361–17365, 2006.

[44] B. Mesko. The Role of Artificial Intelligence in Precision Medicine. *Expert Review of Precision Medicine and Drug Development*, 2(5):239–241, 2017.

[45] M. C. Miller, G. V. Doyle, and L. W. Terstappen. Significance of Circulating Tumor Cells Detected by the CellSearch System in Patients with Metastatic Breast Colorectal and Prostate Cancer. *Journal of Oncology*, 2010:617421, 2010.

[46] J. H. Myung and S. Hong. Microfluidic Devices to Enrich and Isolate Circulating Tumor Cells. *Lab Chip*, 15(24):4500–4511, 2015.

[47] C. G. Rao, T. Bui, M. C. Connelly, G. V. Doyle, I. Karydis, M. R. Middleton, G. Clack, M. Malone, F. A. Coumans, and L. W. Terstappen. Circulating Melanoma Cells and Survival in Metastatic Melanoma. *International Journal of Oncology*, 38(3):755–760, 2011.

[48] A. Regev, S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, H. Clevers, B. Deplancke, I. Dunham, J. Eberwine, R. Eils, W. Enard, A. Farmer, L. Fugger, B. Göttgens, N. Hacohen, M. Haniffa, M. Hemberg, S. Kim, P. Klenerman, A. Kriegstein, E. Lein, S. Linnarsson, E. Lundberg, J. Lundeberg, P. Majumder, J. C. Marioni, M. Merad, M. Mhlanga, M. Nawijn, M. Netea, G. Nolan, D. Pe'er, A. Phillipakis, C. P. Ponting, S. Quake, W. Reik, O. Rozenblatt-Rosen, J. Sanes, R. Satija, T. N. Schumacher, A. Shalek, E. Shapiro, P. Sharma, J. W. Shin, O. Stegle, M. Stratton, M. J. Stubbington, F. J. Theis, M. Uhlen, A. van Oudenaarden, A. Wagner, F. Watt, J. Weissman, B. Wold, R. Xavier, and N. Yosef. The Human Cell Atlas: From Vision to Reality. *eLife*, 550(7677):451, dec 2017.

[49] S. Robertson, H. Azizpour, K. Smith, and J. Hartman. Digital Image Analysis in Breast Pathology - from Image Processing Techniques to Artificial Intelligence. *Translational Research*, 194:19–35, 2018.

[50] M. Saha and C. Chakraborty. HER2Net: A Deep Framework for Semantic Segmentation and Classification of Cell Membranes and Nuclei in Breast Cancer Evaluation. *IEEE Transactions on Image Processing*, 27(5):2189–2200, may 2018.

[51] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv:1312.6034*, 2013.

[52] J. B. Smerage, G. T. Budd, G. V. Doyle, M. Brown, C. Paoletti, M. Muniz, M. C. Miller, M. I. Repollet, D. A. Chianese, M. C. Connelly, L. W. Terstappen, and D. F. Hayes. Monitoring Apoptosis and Bcl-2 on Circulating Tumor Cells in Patients with Metastatic Breast Cancer. *Molecular Oncology*, 7(3):680–692, 2013.

[53] B. W. Stewart and C. P. Wild. *World Cancer Report 2014*. 2014.

[54] J. F. Swennenhuis, A. G. Tibbe, R. Levink, R. C. Sipkema, and L. W. Terstappen. Characterization of Circulating Tumor Cells by Fluorescence In Situ Hybridization. *Cytometry Part A*, 75A(6):520–527, 2009.

[55] A. G. Tibbe, M. Craig Miller, and L. W. Terstappen. Statistical Considerations for Enumeration of Circulating Tumor Cells. *Cytometry Part A*, 71A(3):154–162, 2007.

[56] G. van Dalum, G.-J. Stam, L. F. Scholten, W. J. Mastboom, I. Vermes, A. G. Tibbe, M. R. de Groot, and L. W. Terstappen. Importance of Circulating Tumor Cells in Newly Diagnosed Colorectal Cancer. *International Journal of Oncology*, 46(3):1361–1368, 2015.

[57] M. E. Vandenberghe, M. L. Scott, P. W. Scorer, M. Söderberg, D. Balcerzak, and C. Barker. Relevance of Deep Learning to Facilitate the Diagnosis of HER2 Status in Breast Cancer. *Scientific Reports*, 7:45938, dec 2017.

[58] G. Wang. A Perspective on Deep Imaging. *IEEE Access*, 4:8914–8924, 2016.

[59] F. A. Wolf, P. Angerer, and F. J. Theis. SCANPY: Large-Scale Single-Cell Gene Expression Data Analysis. *Genome Biology*, 19(1):15, dec 2018.

[60] S. Yang, B. Fang, W. Tang, X. Wu, J. Qian, and W. Yang. Faster R-CNN Based Microscopic Cell Detection. In *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pages 345–350. IEEE, dec 2017.

[61] F. Yellin, B. D. Haeffele, and S. Roth. Multi-Cell Detection and Classification Using a Generative Convolutional Model. In *CVPR 2018*, pages 8953–8961, 2018.

[62] M. D. Zeiler and R. Fergus. Visualizing and Understanding Convolutional Networks. In *European Conference on Computer Vision*, pages 818–833. Springer, 2014.

[63] J. Zhao, M. Zhang, Z. Zhou, J. Chu, and F. Cao. Automatic Detection and Classification of Leukocytes Using Convolutional Neural Networks. *Medical & Biological Engineering & Computing*, 55(8):1287–1301, aug 2017.

I

Research Context of Mathematical Imaging and Machine Learning

In the following sections, we will give a very brief introduction to mathematical imaging and machine learning. This introduction is far from complete and will only introduce the most important concepts used throughout this thesis. For more details in mathematical imaging see [30] and for more details on machine learning see [7].

Mathematical Image Analysis

The field of *mathematical imaging* is a rather new multidisciplinary field that arose with the invention of digital imaging. With the huge amount of digital images that is nowadays produced, the importance of image processing is constantly growing. The main applications can be found in biomedical sciences, robotics, satellite imagery, geoscience, material science, but nowadays also in autonomous driving and smartphone apps for image and video processing. The main tasks addressed by image processing are automated image denoising, segmentation, reconstruction, tracking and flow estimation. The subfield of *mathematical image analysis* is focused on both, developing a theoretical foundation why certain processing tools are helpful and also developing new image processing tools based on underlying mathematical models. The image is either seen as a continuous function $u : \Omega \to \mathbb{R}$ with $\Omega \subset \mathbb{R}^d$ or a discrete matrix $U \in \mathbb{R}^{n_1 \times \cdots \times n_d}$. For time-independent RGB color images it is $d = 3$ and for grayscale images it is $d = 2$. The remainder of this section is used to briefly introduce the most important concepts of mathematical imaging we used in our work. We start with an introduction of variational methods for image processing tasks in the continuous setting.

Variational methods Let $u \in \mathcal{U}$ be an image that we want to reconstruct, $K : \mathcal{U} \to \mathcal{V}$ a forward operator mapping from the function space $\mathcal{U}$ to another function space $\mathcal{V}$ and let $f \in \mathcal{V}$ be the function modeling the measured data potentially corrupted by noise. We assume that

$$f \approx Ku \tag{1}$$

resulting in an inverse problem. This inverse problem is typically ill-posed because the stability assumption according to Hadamard is violated due to noisy or incomplete data. The forward operator K models the image formation process - for microscopy data this can be either the identity $K = I$ for direct, undisturbed measurements or a blurring kernel modeling a defocus of the microscope. To solve the inverse problem for u, we minimize a functional J consisting of a data discrepancy term, measuring how well the equality $f = Ku$ is fulfilled, and a regularization functional $R(u)$ imposing some prior constraints on u, i.e.

$$J(u) = D(f, Ku) + \alpha R(u) \to \min_u . \tag{2}$$

The regularization parameter α balances the influence of both terms. The data as well as the regularization term are both adapted to the processing task and the data. The choice of D is normally based on the expected noise in f (see [31] for more details) and the processing task. Common data-discrepancy terms are the L^p-norms $||Ku - f||_{L^p}$ especially for $p = 1$ and $p = 2$. The regularization functional is chosen based on the prior information that should be incorporated in the model. Common choices are piecewise constancy (TV regularization), smoothness (L^2-norm regularization of the image gradient) or sparsity (L^1-norm regularization). To minimize the functional J it requires some form of derivative of J. If J is not differentiable in the classical sense, advanced optimization algorithms are needed to solve (2).

In the following paragraph, we will present a variational model for image segmentation that we later use in Chapter 1.

Segmentation using active contour methods In mathematical imaging the problem of segmentation describes the task to automatically detect regions of interest in an image u. In biomedical imaging, segmentation is often used either as a preprocessing step for feature extraction or tracking, or alternatively as a post-processing step after image reconstruction to detect regions of interest in the reconstructed image. While simple segmentation models like histogram thresholding often fail for experimental datasets, nonlinear variational methods offer great potential. In general there are two ways to define a region - either we describe the edge or the interior. In the following we will describe an instance of the class of region-based segmentation models.

In 1989 Mumford and Shah introduced the following variational model for image segmentation [24]

$$J^{\mathrm{MS}}(u, C) = \int_\Omega |f(x) - u(x)|^2 \mathrm{d}x + \alpha \cdot \mathrm{Length}(C) + \beta \cdot \int_{\Omega \setminus C} |\nabla u(x)|^2 \mathrm{d}x \longrightarrow \min_{u,C} . \tag{3}$$

Here u is a differentiable function that is allowed to be discontinuous on the contour C defining the segmentation. This enables that the solution u can contain sharp edges between image regions. Apart from the contour, the L^2- norm regularization of the gradient

Figure 3: *Visualization of a contour moving towards the final segmentation.*

enforces u to be smooth. This model was adapted by Chan and Vese in 2001 [8]. They restrict u to have only two image regions Ω_1 and Ω_2 both having constant intensity values c_1 and c_2 respectively. Again, C describes the contour separating both regions and thereby defining the segmentation. This can be formalized as

$$J^{\mathrm{CV}}(c_1, c_2, C) = \lambda_1 \cdot \int_{\Omega_1} (f(x) - c_1)^2 \mathrm{d}x + \lambda_2 \cdot \int_{\Omega_2} (f(x) - c_2)^2 \mathrm{d}x + \alpha \cdot \mathrm{Length}(C) + \beta \cdot \mathrm{Area}(\Omega_1) \quad (4)$$

where $\lambda_1, \lambda_2, \alpha, \beta$ are constants weighting the corresponding term. c_1 and c_2 are constants that need to be determined during the minimization of $J^{\mathrm{CV}}(c_1, c_2, C)$. This model is called the "active contour without edges" model. During the minimization of (4) the contour C evolves until it reaches a minimum which, in the ideal case, describes the object of interest's boundaries. A visualization of this process is shown in Figure 3.

Total Variation regularization As mentioned before the regularization functional can be used to incorporate prior knowledge about the solution u in the variational model. A common assumption in case of noisy data is that the underlying image $u \in \mathcal{U}$ is almost everywhere smooth. Linear filtering methods that penalize for example the L^2-norm of the gradient therefore assume that the image contains no sharp edges. Yet, in most images sharp edges are needed to describe for example object's boundaries. Therefore Rudin, Osher and Fatemi introduced the concept of nonlinear Total Variation (TV) regularization for image processing [29]. The TV functional is defined as

$$TV(u) := \sup_{\substack{g \in C_0^\infty(\Omega; \mathbb{R}^2) \\ ||g||_\infty < 1}} \int_\Omega u \nabla \cdot g \, \mathrm{d}\mu = \int_\Omega |Du| \quad (5)$$

where Du denotes the gradient of u in a distributional sense. The corresponding space of functions with bounded variation is defined as $BV(\Omega) := \{u \in L^1(\Omega) | TV(u) < \infty\}$. For functions u contained in the Sobolev space $W^{1,1}$ it holds that

$$TV(u) = ||\nabla u||_{L^1}. \quad (6)$$

This motivates the intuitive understanding of TV penalizing jumps in u and enforcing a sparse (i.e. zero almost everywhere) gradient. Small oscillation will be smoothed but large gradient

values occurring at edges are retained in the solution. This results in piecewise constant, cartoon-like images. This directly motivates the use of TV regularization in the aforementioned segmentation model (4) where we aim to find a piecewise constant image with only two intensity values. By using the co-area-formula it can even be shown that the total variation of a characteristic

$$u(x) = 1_C(x) = \begin{cases} 1 & \text{if } x \in \Omega_1 \cup C \\ 0 & \text{if } x \in \Omega_2 \end{cases}$$

corresponds exactly to the contour length $\text{Length}(C)$ in (4).

The minimization of the total variation with the steepest descent method results in the following partial differential equation that is called TV-flow [2] and is used in Chapter 1:

$$\begin{aligned} u_t(t,x) &= -\nabla \cdot \left(\frac{Du}{|Du|}\right) && \text{in } \ (0,\infty)\times\Omega \\ u(0,x) &= f(x) && \text{for } \ x \in \Omega. \end{aligned} \tag{7}$$

Here, Du is defined as above, $\nabla\cdot$ denotes the divergence and we assume Neumann boundary conditions. For $f(x) = 1_C(x)$ the unique solution of (7) according to [5] is

$$u(t,x) = (1-\lambda_C t)^+ 1_C(x) \quad \text{with} \quad \lambda_C = \frac{\text{Length}(C)}{\text{Area}(C)}.$$

An intuitive interpretation why the TV regularizer leads to cartoon-like images can be seen in (7). If x lies in a rather smooth area, $|Du|$ is small and $\nabla\cdot\left(\frac{Du}{|Du|}\right)$ will smooth rather strongly. On the other hand, if x lies on an edge, $|Du|$ is large and will result in only little to no smoothing. Therefore, minimizing the TV functional leads to a *nonlinear* diffusion process.

Machine Learning

The purpose of machine learning (ML), a subfield of artificial intelligence (AI), is to develop algorithms that learn to make predictions from input data without being explicitly programmed how to make these predictions. ML algorithms use statistical analysis for their predictions which should improve when new input data, the algorithm can "learn" from, becomes available. In general, we can distinguish between *supervised* and *unsupervised* methods.

For supervised problems, we require as input data a set S of N instances $x^i \in \mathbb{R}^n$ with required output data $y_i \in \mathbb{R}^l$, so $S = \{(x^i, y^i) \in \mathbb{R}^n \times \mathbb{R}^l, i = 1,\dots,N\}$. In case of a classification problem, y_i would be the label that was given to input x^i. For supervised learning, the set of available examples S is split into a so called training set S_{train} that the algorithm uses to learn making the correct predictions and a test set S_{test} to evaluate the algorithm's performance on examples it has not seen before. If $f : \mathbb{R}^n \to \mathbb{R}^l$ is the underlying (unknown) function f mapping input to output data the goal is to approximate f in an optimal way by only knowing the instances $f(x^i) = y^i$. To decide which mapping is "optimal" a loss function L similar to J in (2) is defined. The loss function depends on the mapping f and takes values close to zero if f is a close approximation of the true underlying mapping between input and

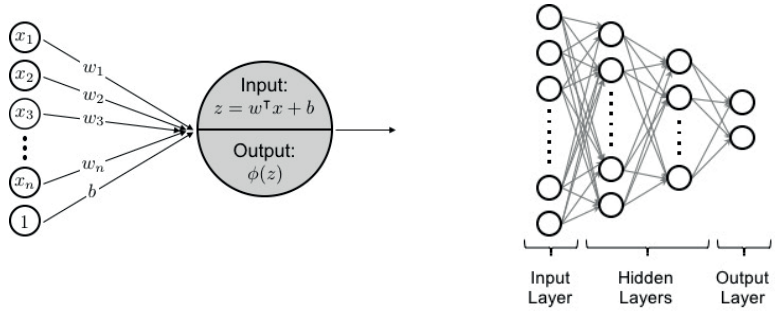
output data. Therefore, in machine learning problems, the following minimization problem is solved

$$J(f) = L(f) + \alpha R(f) = \sum_{i=1}^{N} L(f(x^i), y^i) + \alpha R(f) \rightarrow \min_f . \tag{8}$$

As in (2), $R(f)$ is a regularization functional incorporating prior information about f and α weights its influence on the solution f. Although this minimization problem is very similar to the minimization problem in the previous section, the fundamental difference is that the mapping K in (2) is known and we search for an optimal u but for a machine learning problem (8), the input data x is known but we search for an optimal mapping f between input and output data. The most common task in supervised learning is classification, where we try to map x^i to a (discrete) label y^i. For cell analysis, this could be a mapping of input features x^i to a cell class. Examples for machine learning methods used for classification are Support Vector Machines [10], Decision Trees [26] and Neural Networks which are described below.

For unsupervised machine learning, we only have input data x^i but no required outputs, i.e. $S = \left\{x^i \in \mathbb{R}^n, i = 1, \ldots, N\right\}$. In this case, the objective is to find hidden patterns in the input data that describe the structure of different inputs. One class of unsupervised learning methods are clustering methods, where the aim is to find groups of x^i that share common characteristics which separate them from the other examples. Well-known examples for clustering algorithms are for example the K-Means algorithm [23], Nearest Neighbors approaches [28] and Spectral Clustering [25]. Another group of unsupervised learning algorithms consists of dimensionality reduction methods, where the aim is to find a representation of the input data in a lower dimensional space. Prominent examples are Principle Component Analysis [18], Singular Value Decomposition [12] and again Neural Networks.

Preprocessing of images as input data For applications of machine learning on imaging data, we can distinguish two different classes of input data. Traditionally, the learning algorithms (both supervised and unsupervised) do not work directly on the image itself but on "features" that were extracted before. That means, we transform the original image $x \in \mathbb{R}^n$ into the feature space $\mathbb{R}^m$ (m< n) using a feature extraction function $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$. The learning algorithm afterwards uses the feature vector $F(x)$ as an input instead of the image x itself. In biomedical applications the feature space often consists of hand-crafted, intuitive features like edges or physical properties like roundness, size or intensity. In that way, we already define what we think are the most informative features the algorithm needs, to perform the learning task. A disadvantage is that we already bias the algorithm by using hand-crafted features. A hidden feature, which is not directly related to physical or image properties, might be missed in that way. Recent advances in Neural Networks (NN) are therefore using the original image x as their input. The methods consist of two parts - a feature-extraction part that is also learned from the input data and a task-specific function. That means the algorithm learns itself which features are most useful to extract for the desired task and is therefore not biased by the human interpretability of input features.



(a) Artificial Neuron (b) Neural Network

Figure 4: *Sketch of an artificial neuron (a) and an example for a neural network with two hidden layers (b).*

Neural Networks and Deep Learning Neural Networks are graph structures where each node is an artificial neuron which is inspired by neurons in the brain. An artificial neuron with input $x \in \mathbb{R}^n$ is defined as

$$\eta(x; \phi, w, b) = \phi(w^\intercal x + b) \tag{9}$$

where w is the weight vector, b is the bias and ϕ is the activation function. Thus, the neuron receives $x_1, \ldots, x_n$ inputs, assigns a weight w_i to each of them and sums up the weighted inputs. Moreover, a bias b is added to the sum $w^\intercal x$. Finally, the neuron outputs a (nonlinear) activation function ϕ applied to the so called scores $w^\intercal x + b$. A sketch of the mechanism of an artificial neuron is presented in Figure 4 (a). A combination of multiple neurons in an acyclic, directed graph is called a *Neural Network* (see Figure 4 (b)). The first layer of neurons is called *input layer*, the last layer is the *output layer* and all layers in between are so called *hidden layers*. A network that has no or only very few hidden layers is called a *shallow network* while a network with several hidden layers is called *deep*. Neural networks are a class of mappings f from input to output data. By solving the minimization problem (8) all weights and biases are optimized with respect to the input data and the desired output.

As mentioned before, minimizing a functional $J(f)$ requires the derivative with respect to the parameters to optimize, i.e. $\frac{\partial J}{\partial f}$. Since f itself is a composition of stacked layers, the chain rule can be used to determine the derivative. The gradients with respect to the parameters in layer $(k-1)$ can be expressed in terms of the gradients with respect to parameters in the following layer k (see [16] for a more detailed description). To compute the derivative, we therefore start to compute the derivative of the loss function with respect to the parameters of the last layer and then propagate backwards through all layers. This algorithm is called the *back-propagation algorithm*. It can be seen as the opposite mapping to the forward path through the network, which maps x to the output y. The back-propagation algorithm therefore allows to efficiently train also more complex networks as long as the activation functions are differentiable so that all partial derivatives can be computed.

The key element for the flexibility of neural networks are the nonlinear activation functions ϕ. They allow to approximate arbitrary functions to any given accuracy. In real world problems this is crucial, since most underlying functions, mapping real word inputs to the desired output, require advanced nonlinearities.

As shown in the example of Figure 4 (b) neural networks are stacked layers of artificial neurons. The neurons in each layer (apart from the input and output layer) are connected to each neuron in the layer before and after. This architecture is called a *fully connected layer*. More formally, for N_{k-1} neurons in the $(k-1)$-th layer and N_k neurons in layer k we define the fully-connected layer I_k^{full} as the function $I_k^{\text{full}} : \mathbb{R}^{N_{k-1}} \to \mathbb{R}^{N_k}$ defined as

$$I_k^{\text{full}}(x; W, b) := \phi(Wx + b). \tag{10}$$

Here $\phi = [\phi_1, \phi_2, \ldots, \phi_{N_k}]$ is a point-wise activation function, $W = [w_1, w_2, \ldots, w_{N_k}] \in \mathbb{R}^{N_k \times N_{k-1}}$ is the weight matrix storing the weights for each neuron in I_k^{full} and $b \in \mathbb{R}^{N_k}$ is the bias vector. For simplicity, we summarize all network parameters that are optimized in the training phase (i.e. all weight matrices W and all biases b) in one variable θ.

Multi-layer network mappings can be separated into two different functions, the feature-extraction function and a task-specific function (see Figure 9). The first part, defined as F, is responsible for extracting optimal features from the input data for the desired task and thereby replaces the manual feature extraction described above. The extracted features are then input to a second, task-specific, function $T : \mathbb{R}^m \to \mathbb{R}^l$. In case of a supervised classification problem, this is a nonlinear classifier where the number of neurons in the last layer corresponds to the number of classes. The function $f : \mathbb{R}^n \to \mathbb{R}^l$ mapping the input x to the desired output y can therefore be described as

$$f(x; \theta) = T\left(F\left(x; \theta_F\right); \theta_T\right). \tag{11}$$

The functions F and T are parametrized by θ_F and θ_T, that means the weights and biases of all layers used in F or T respectively are summarized in the corresponding θ.

Once a deep network is built, the task is to find the optimal weights given the input and output data. This requires the solution of the minimization problem (8) with f defined as in (11). With the growing complexity and depth of network architectures the optimization process became so challenging that efficient training of deep neural networks became a research field on its own, called *Deep Learning* [20]. Although technically the back-propagation algorithm can be used to train every deep network with differentiable activation functions, several practical challenges can arise. A major drawback are vanishing or exploding gradients that prevent the efficient training of early layers. The back-propagation algorithm is based on the chain-rule where several derivatives are multiplied. If all derivatives are smaller than one, this product becomes very small for early layers in deep networks so that these layer's weights are only marginally updated. Alternatively, if all derivatives are larger than one, the product of all partial derivatives becomes very large resulting in severe changes in early layers and an unstable training. Techniques to prevent both cases include new network architectures called *residual networks* [14], adapted activation functions [11] or regularization.

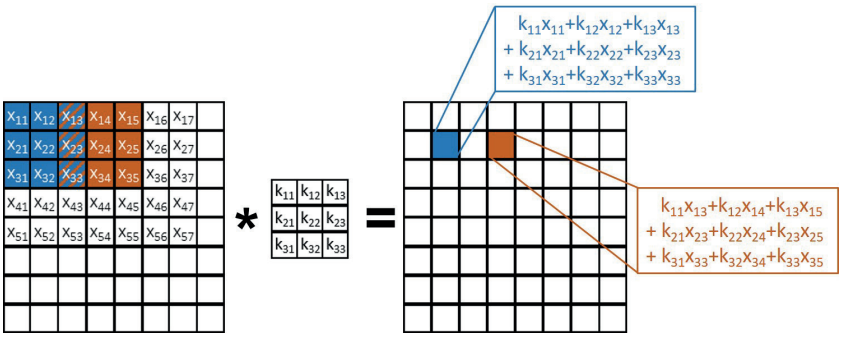


Figure 5: *Visualization of a convolution in a neural network.* Each pixel is connected to all surrounding pixels covered by the convolution kernel.

The choice of the loss function L, that is used for the optimization of the neural network parameters, is strongly related to the addressed task. As in (2), a common choice is the L^2-distance

$$L^i\left(f(x^i;\theta), y^i\right) = ||f(x^i;\theta) - y^i||_{L^2}.$$

For classification networks, we often use distance measures comparing two probability distributions. Examples are the cross-entropy, the Kullback-Leibler divergence and Wasserstein loss functions [3, 33]. Applications of Deep Learning can be found for example in Speech Recognition [17], Robotics [21], Biomedicine [22] and Finance [15]. For an overview of neural network architectures and learning rules see [13], for a survey focusing on neural networks for classification see [36] and for a recent survey on Deep Learning in Neural Networks see [32].

Convolutional Neural Networks In the field of image processing where $x \in \mathbb{R}^{n_1 \times \cdots \times n_d}$ the use of fully connected networks presented in the previous paragraph is often hindered by the huge amount of parameters needed. Assume we have a grayscale image with $100 \times 100 = 10,000$ pixels, a shallow network with only 2 hidden layers of 200 neurons each and an output layer with 2 nodes. This results in $10000 \times 200 + 200 \times 200 + 200 \times 2 = 2,040,400$ weights plus one bias per node to optimize. If the image is an RGB color image instead of grayscale, the number of weights directly increases to $6,040,400$. This shows that for reasonably large images, deeper networks are computationally not feasible and a huge amount of training data is needed to optimize all parameters.

One way to circumvent this problem is to make use of the repetitive structure of images and use a so called *convolutional neural network*. Instead of connecting each pixel in one layer to each pixel in the following layer, a pixel is only connected to the pixels surrounding it through a convolutional (see Figure 5). The number of connected pixels is hereby determined by the kernel size s. Instead of learning all weights of a fully connected layer, only the weights of the convolutional kernels are learned. The same kernel then strides over the image and the resulting convolution plus the bias is input to the activation function. A convolutional layer corresponds to a fully-connected layer with a sparse weight matrix

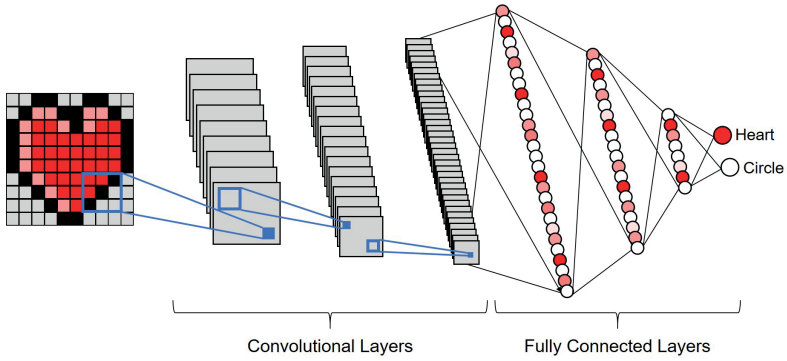


Figure 6: *Visualization of a convolutional neural network for binary classification.* The network consists of three convolutional blocks with increasing number of kernels followed by three fully-connected layers with decreasing number of neurons and the output layer.

where only $(s^2 - 1)$ diagonals are non-zero. Moreover, for each non-zero diagonal all entries have the same value, reflecting that the *same* convolutional strides over the full image. Yet, different to a fully-connected layer, in each layer not only one kernel, i.e. one sparse weight matrix, is learned but a fixed number of N_k kernels. This number is commonly increased with increasing depth of the network. Each kernel results in a new representation of the image and in the subsequent layer these representations are concatenated in the third dimension. Formally, a convolutional layer can be written as the function

$$I_k^{\text{conv}}(x; K, b) := [\phi(z_1), \ldots, \phi(z_{N_k})] \tag{12}$$

where each feature map z_j $(j = 1, \ldots, N_k)$ is defined as

$$z_j(x; K_j) = K_j * x + b_j \tag{13}$$

where $*$ denotes the discrete convolution of x with kernel K_j and b_j the bias. A simple convolutional neural network is visualized in Figure 6.

One way to understand how a CNN is looking at the input data is to visualize the learned features of the network. Figure 7 shows an example of a feature visualization taken from [35]. Here, they visualized the features learned by a convolutional neural network trained on the ImageNet dataset (http://image-net.org). For a fixed feature map, they showed the nine most prominent activations (on the left) and next to it on the right the image patch that activated the feature map. We can see that for each feature map, the nine images are very similar showing for example all a striped pattern (layer two, row one, column two) or all a car wheel (layer three, row two, column two). This reflects that the learned kernels in a convolutional neural network highlight on very specific image features. Moreover, we see that the scale of the features increases with the layer's depth. In the first layer, only very small scale features like edges are learned while in the third layer the feature map already highlights on much more complex patterns (row one, column one) or on the upper body and head of a person (row three, column three). This shows the *hierarchical feature*

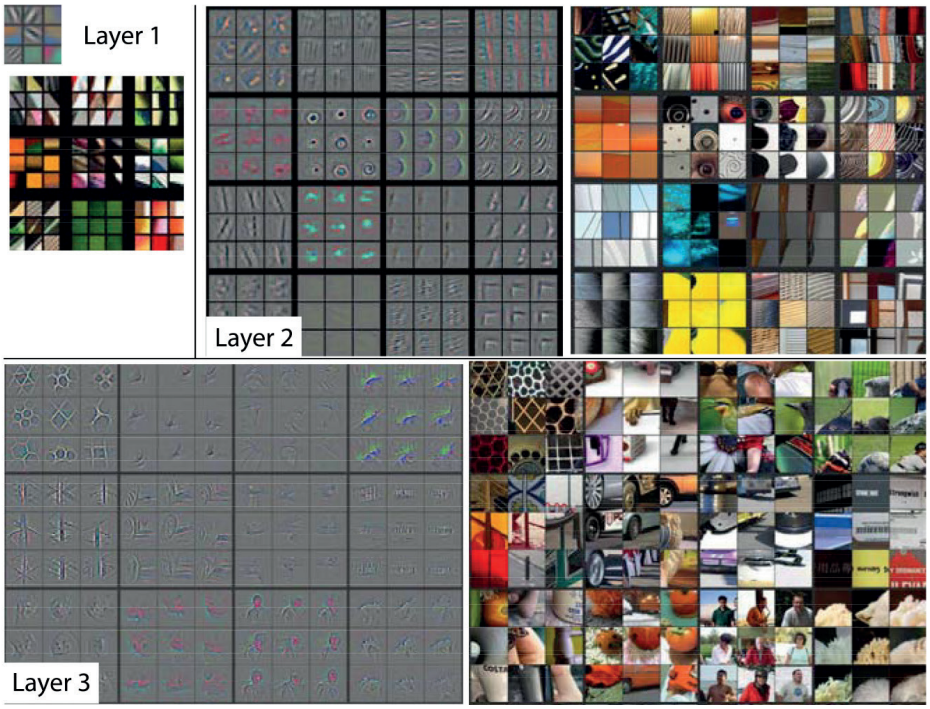


Figure 7: *Visualization of features in a fully trained network taken from [35].* A comparison of the three different layers shows that the scale of features increases for deeper layers.

extraction property of a convolutional network. An intuitive idea why deeper layers learn larger scale features of the input image can be seen in Figure 8. We choose one pixel in layer k (highlighted in blue) and a kernel size of 3×3. If we then look in the previous layer $(k-1)$ which pixels influenced the activation of the chosen pixel in layer k, we see that it is a 3×3 window around the original pixel. If we then look even one layer before, i.e. in layer $(k-2)$, it is already a window of 5×5 pixels that influence the activation of one pixel in layer k. The field of pixels influencing the activation (marked for each layer in blue in Figure 8) is called the *receptive field*. We can see that it increases with every layer so the deeper the layer, the larger is the receptive field in the input image. Therefore, deeper layers learn large scale features of the input image and early layers have a very localized view on the input image and therefore learn small scale features.

Autoencoders In case of unsupervised learning, no output labels $y^i \in \mathbb{R}^l$ are available. Yet, to find structure in the data, we can map our input data x^i to x^i itself but we restrict the mapping function $f(x;\theta)$ so that it learns more than a simple identity mapping. This neural network architecture is called an *autoencoder* [4]. Since autoencoders map $f : \mathbb{R}^n \to \mathbb{R}^n$ (or in case of images $f : \mathbb{R}^{n_1 \times \cdots \times n_d} \to \mathbb{R}^{n_1 \times \cdots \times n_d}$) they have a significantly different structure than classification networks that map $x^i \in \mathbb{R}^n$ to $y^i \in \mathbb{R}^l$ with $l << n$. A comparison of both

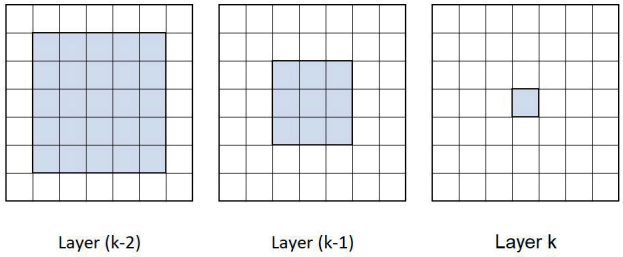
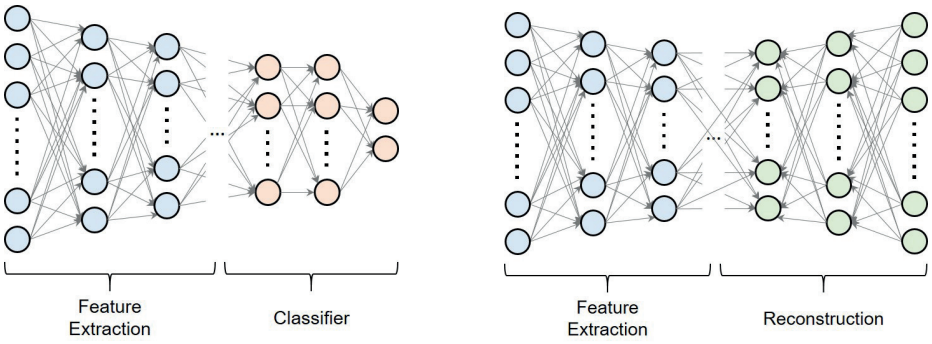


Figure 8: *Visualization of the receptive field of three subsequent layers.*



(a) Classification NN (b) Autoencoder

Figure 9: *Example of a neural network for classification (a) and an autoencoder (b) with the same encoding architecture.*

architectures in case of a fully-connected network is shown in Figure 9. For autoencoders, the task-specific function $T : \mathbb{R}^m \to \mathbb{R}^n$ is a reconstruction function that maps the features $F(x;\theta_F)$ back to the input space. In order to compress the input data and learn the most important features needed to reconstruct the input as best as possible, the feature space $\mathbb{R}^m$ is most of the time lower dimensional, i.e. $m < n$. Yet, similar to over-complete dictionaries in dictionary learning, there are also approaches where a higher dimensional feature space is chosen. We refer to the feature space as the *latent space* and call $F(x;\theta_F) \in \mathbb{R}^m$ the *latent variables*. In Chapter 6, we investigate the structural difference of the latent space for the different network architectures. Although the encoding and decoding architectures for autoencoders can be chosen independently, it is common to use symmetric autoencoders. Similar network architectures are also successfully used for image processing tasks like denoising [34] and segmentation [27].

Relation of neural networks to filtering and variational methods In the previous sections, we have already seen that variational methods and deep learning methods are both energy minimization problems with a similar structure. The nonlinear diffusion equation (7) arising from the minimization of the total variation (6) can be interpreted as a

forward-path through a convolutional layer with fixed weights and a fixed activation function. This motivates the question if the total variation regularization is really the best regularization for the given data or if other weights and activation functions would improve the results. This question is addressed by a class of neural network architectures called *variational networks* [9, 19]. The idea is to replace the fixed regularization functional of the variational method (2) with a trainable convolutional neural network. Variational networks therefore combine the idea of both fields and result in a bi-level optimization problem where we minimize for both, the solution u and the network parameters θ. Another idea of relating regularization functionals with neural networks follows from the ideas presented in [1, 6]. It can be shown that there exists an energy functional E for which one step in a gradient scheme to minimize E corresponds to a forward path through an autoencoder with tied-weights (i.e. an autoencoder where the decoding weight matrix is the transpose of the encoding weight matrix). This again shows the close relation between optimal regularization functionals and neural networks.

Bibliography

[1] G. Alain and Y. Bengio. What Regularized Auto-Encoders Learn from the Data-Generating Distribution. *Journal of Machine Learning Research*, 15(1):3563–3593, 2012.

[2] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing Total Variation Flow. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 331(11):867–872, 2000.

[3] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein Generative Adversarial Network. In *International Conference on Machine Learning*, pages 214–223, 2017.

[4] P. Baldi. Autoencoders, Unsupervised Learning, and Deep Architectures. In *ICML Unsupervised and Transfer Learning*, pages 37–49, 2012.

[5] G. Bellettini, V. Caselles, and M. Novaga. The Total Variation Flow in RN. *Journal of Differential Equations*, 184(2):475–525, 2002.

[6] Y. Bengio, A. Courville, and P. Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.

[7] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[8] T. F. Chan and L. A. Vese. Active Contours Without Edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[9] Y. Chen, W. Yu, and T. Pock. On Learning Optimized Reaction Diffusion Processes for Effective Image Restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5261–5269, 2015.

[10] C. Cortes and V. Vapnik. Support-Vector Networks. *Machine Learning*, 20(3):273–297, 1995.

[11] X. Glorot, A. Bordes, and Y. Bengio. Deep Sparse Rectifier Neural Networks. In *International Conference on Artificial Intelligence and Statistics*, pages 315–323, 2011.

[12] G. H. Golub and C. Reinsch. Singular Value Decomposition and Least Squares Solutions. *Numerische Mathematik*, 14(5):403–420, 1970.

[13] M. T. Hagan, H. B. Demuth, and M. H. Beale. *Neural Network Design*. Martin Hagan, 2014.

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[15] J. B. Heaton, N. G. Polson, and J. H. Witte. Deep Learning for Finance: Deep Portfolios. *Applied Stochastic Models in Business and Industry*, 33(1):3–12, 2017.

[16] C. F. Higham and D. J. Higham. Deep Learning: An Introduction for Applied Mathematicians. *arXiv:1801.05894*, 2018.

[17] G. E. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury. Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[18] I. T. Jolliffe. Principal Component Analysis. In *International Encyclopedia of Statistical Science*, pages 1094–096. 2011.

[19] E. Kobler, T. Klatzer, K. Hammernik, and T. Pock. Variational Networks: Connecting Variational Methods and Deep Learning. In *Pattern Recognition*, pages 281–293, 2017.

[20] Y. A. LeCun, Y. Bengio, and G. E. Hinton. Deep Learning. *Nature*, 521(7553):436, 2015.

[21] I. Lenz, H. Lee, and A. Saxena. Deep Learning for Detecting Robotic Grasps. *International Journal of Robotics Research*, 34(4-5):705–724, 2015.

[22] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. Arindra Adiyoso Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sanchez. A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, 42:60–88, 2017.

[23] J. Macqueen. Some Methods for Classification and Analysis of Multivariate Observations. In *Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[24] D. Mumford and J. Shah. Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[25] A. Y. Ng, M. I. Jordan, and Y. Weiss. On Spectral Clustering: Analysis and an Algorithm. In *Advances in Neural Information Processing Systems*, pages 849–856, 2002.

[26] J. R. Quinlan. Induction of Decision Trees. *Machine Learning*, 1(1):81–106, 1986.

[27] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[28] S. T. Roweis and L. K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, 2000.

[29] L. I. Rudin, S. J. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[30] O. Scherzer. *Handbook of Mathematical Methods in Imaging*. Springer Science+ Business Media, 2010.

[31] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, F. Lenzen, M. Grasmair, and M. Haltmeier. *Variational Methods in Imaging*. Springer Science+ Business Media, 2009.

[32] J. Schmidhuber. Deep Learning in Neural Networks: An Overview. *Neural Networks*, 61:85–117, 2015.

[33] R. Vidal, J. Bruna, R. Giryes, and S. Soatto. Mathematics of Deep Learning. *arXiv:1712.04741*, 2017.

[34] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research*, 11(Dec):3371–3408, 2010.

[35] M. D. Zeiler and R. Fergus. Visualizing and Understanding Convolutional Networks. In *European Conference on Computer Vision*, pages 818–833. Springer, 2014.

[36] G. P. Zhang. Neural Networks for Classification: a Survey. *IEEE Transactions on Systems, Man and Cybernetics*, 30(4):451–462, 2000.

RC

Chapter 1

Multiscale Segmentation via Bregman Distances and Nonlinear Spectral Analysis

Leonie L. Zeune, Guus van Dalum, Leon W.M.M. Terstappen,
Stephan A. van Gils and Christoph Brune

SIAM journal on imaging sciences 10.1 (2017): 111-146

Abstract

In biomedical imaging reliable segmentation of objects (e.g. from small cells up to large organs) is of fundamental importance for automated medical diagnosis. New approaches for multi-scale segmentation can considerably improve performance in case of natural variations in intensity, size and shape. This paper aims at segmenting objects of interest based on shape contours and automatically finding multiple objects with different scales. The overall strategy of this work is to combine nonlinear segmentation with scales spaces and spectral decompositions. We generalize a variational segmentation model based on total variation using Bregman distances to construct an inverse scale space. This offers the new model to be accomplished by a scale analysis approach based on a spectral decomposition of the total variation. As a result we obtain a very efficient, (nearly) parameter-free multiscale segmentation method that comes with an adaptive regularization parameter choice. To address the variety of shapes and scales present in biomedical imaging we analyze synthetic cases clarifying the role of scale and the relationship of Wulff shapes and eigenfunctions. To underline the potential of our approach and to show its wide applicability we address three different experimental biomedical applications. In particular, we demonstrate the added

(a) Sizes (b) Intensities (c) Shapes

Figure 1.1: *Minimal examples where multiscale image segmentation could offer added value.*

benefit for identifying and classifying circulating tumor cells and present interesting results for network analysis in retina imaging. Due to the nature of nonlinear diffusion underlying, the mathematical concepts in this work offer promising extensions to nonlocal classification problems.

1.1 Introduction

In mathematical imaging the problem of segmentation refers to the process of automatically detecting regions, objects or patterns of interest. This is of particular importance in biomedical imaging for cell or organ quantification as well as in materials science and engineering. The main goal of this work is to present a new multiscale segmentation method combining variational inverse scale-space methods with nonlinear spectral analysis.

Image segmentation. The task of image segmentation can sometimes be addressed by simple methods which are solely based on intensity or histogram thresholds. However those methods quickly fail in more complex experimental scenarios, where challenges in terms of different sizes, intensities, contrasts or uncertainties like noise occur. Thus, a class of commonly used mathematical techniques to overcome some of the difficulties are nonlinear variational methods. In general there are two different ways of describing a region, either by its edge or by the interior of the region. Therefore two main concepts of addressing segmentation by an energy minimization problem evolved in the previous decades: edge-based and region-based segmentation. While edge-based segmentation separates regions based on discontinuity information [19, 39, 40], region-based segmentation separates by similarity measures within the regions. In this paper we focus on a region-based approach.

The proposed method in this work is mainly built upon the Chan-Vese (CV) model [24]. For a given image function $f : \Omega \to \mathbb{R}$ the domain $\Omega \subset \mathbb{R}^d$ should be separated into two regions Ω_1 and $\Omega_2 := \Omega \setminus \Omega_1$ via the following variational energy minimization method

$$\int_{\Omega_1} (f(x) - c_1)^2 \mathrm{d}x + \int_{\Omega_2} (f(x) - c_2)^2 \mathrm{d}x + \alpha \cdot Per(C) \longrightarrow \min_{C, c_1, c_2} \tag{1.1}$$

where C is the desired contour separating Ω_1 and Ω_2 and c_i stands for a desired average intensity value within region Ω_i. $Per(C)$ denotes the length of the contour C separating Ω_1 and Ω_2 and can be controlled via a regularization parameter $\alpha > 0$. An easy example is the image f given in Figure 1.1 (a) where the domain Ω can be segmented with respect

to foreground and background. In contrast to simple histogram thresholding methods this holds true even in the case of strong noise if the parameter α is chosen adequately. A generalization of this model was introduced in [60] and in [23] the model has also been extended to vector-valued images like color images. Recently in [67] an extension of the CV model that can handle artifacts and illumination bias in images has been proposed.

Challenges/Questions. The CV model is very useful in segmenting regions of interest which have very similar intensity values, e.g. Figure 1.1 (a). However, automatically detecting single objects based on their size is more challenging. Even with a varying parameter α controlling the contour length (forward scale-space), it is for example not possible to detect the smallest object as a singleton. A similar challenge occurs when segmenting separate objects due to their intensity values, e.g. Figure 1.1 (b). Increasing the number of constants c_i to four is suboptimal because we usually a-priori do not know the number of objects. Varying a threshold or parameter α could lead to a correct segmentation but the estimated intensity constants c_1, c_2 will likely be incorrect. *Hence, is it possible to automatically detect multiple scales in a nonlinear variational image segmentation model, for instance with respect to different object sizes or object intensities? Can the segmentation of an image automatically be decomposed with respect to those scales?*

Many region-based segmentation methods only use constraints on the contour length or curvature as regularization. However, in view of shape optimization and dictionary learning an approach that could also automatically separate objects with respect to their shape (cf. Figure 1.1 (c)) would be very interesting. *Hence, what is the role of geometric shapes in a multiscale segmentation approach?*

Scale-space methods. In the previous decades there has been a continuous interest in the analysis of different scales and the construction of scale spaces in imaging. In general it is desired to automatically detect all scales present in an image and simultaneously determine which scales are informative and contribute most to the image. For segmentation this problem is addressed in the fundamental works by Witkin and Koenderink [41, 62]. In relation to those, several methods to detect and analyze interesting scales have been proposed, see for example [31, 42, 43, 44, 59, 61]. The underlying scale-space that is examined is defined by a linear diffusion process. A drawback of those approaches is that linear diffusion smoothes edge information and is therefore in general not suitable for applications where one is interested in retaining sharp edge information. Especially in biomedical image applications this is often the case. Therefore those theories were extended to non-linear diffusion processes, see [27, 48, 49]. A drawback of these approaches is that their analysis of scales is not fully automatic and can only be used in a forward approach, thus going from fine to coarse scales and then trying to find a backward relation. In this work we concentrate on nonlinear diffusion processes for segmentation where scale automatically relates to intensity and size.

A prominent example of a variational method for nonlinear diffusion is the ROF model [50]. With increasing regularization parameter α a sequence of functionals generates a nonlinear forward scale space flow that filters signals from fine to coarse. However in this process the total variation regularization functional is known to lead to a systematic contrast loss in the filtered image u [46], whereas the main discontinuities in the signal remain at

their position in the domain. To tackle the problem of intensity loss Osher et al. proposed in [50] an iterative contrast enhancement procedure based on Bregman distances. This approach is known to generate a nonlinear inverse scale space flow generating filtered signals from coarse to fine and with improved quality. This idea was successfully applied to more general inverse problems [9, 13, 16].

Spectral methods. Recently Gilboa [32, 33] developed a framework to detect scales based on the nonlinear total variation diffusion process. The total variation is known to retain edge information while smoothing the signal apart from the edges. In this framework scales are detected based on a spectral decomposition of the given image into TV eigenfunctions [46]. This concept does not only hold true for higher-order regularization functionals [9, 52, 54] but more generally for convex, one-homogenous functionals J with corresponding nonlinear eigenvalue equations [15, 17] of the following form

$$\lambda u \in \partial J(u)$$

where $\partial J(\cdot)$ denotes the subdifferential of J. When minimizing for instance the total variation J, those eigenfunctions u simply loose contrast whereas the overall structure of the function remains the same. The magnitude of contrast loss is related to the eigenvalue λ. The eigenfunctions shape is determined by the chosen norm in the TV functional which can be adapted to the application of interest. In this way signals are not linearly smoothed to overcome scales but are step-by-step transformed to a composition of nonlinear eigenfunctions at coarser scales. A spectral response function can be used to examine which scales have a strong contribution to the original signal and to design filters of certain scales. Moreover, such spectral method can be combined with forward and inverse scale space approaches [17].

Main contribution. In this paper we will extend the idea of (inverse) scale space methods known for nonlinear diffusion processes to segmentation and shape detection problems. Our main goals are:

1. A novel inverse multiscale segmentation method based on Bregman iterations
2. An adaptive regularization parameter strategy for α (independent of c_1, c_2)
3. A spectral analysis for segmentation regarding shapes of eigenfunctions

A TV-based forward scale space for segmentation can easily be derived from the CV model via an increasing regularization parameter. We extend this framework by an inverse scale space for segmentation, still based on the CV model and therefore on a nonlinear diffusion process. For this purpose, we make use of Bregman iterations, among others well-known for improving total variation denoising results. The relation between the forward scale space and the inverse scale space is examined. Both iterative strategies are accomplished by a spectral transform and response function, which are used to easily examine scales and to filter certain scales. Since our method uses the total variation as a nonlinear diffusion process, we can make use of relatively easy and fast numerical and parallel implementation schemes developed in recents years.

Organization. This work is organized as follows. In Section 1.2.1 we start with a revision of the segmentation model by Chan and Vese including its convexification. Together with a revision of the iterative denoising strategy using Bregman iterations in Section 1.2.2, the combination of those two concepts forms the first ingredient of our new inverse scale space method for multiscale segmentation introduced in Section 1.2.3. An interesting interpretation of the method as an adaptive regularization method is presented in Section 1.2.4. Section 1.3 deals with nonlinear spectral methods and contains the second main ingredient of our new approach. In 1.3.1 we start with a brief summary of generalized nonlinear total variation functionals addressing different eigenshapes and continue in 1.3.2 with recent works by Gilboa et al. solving related nonlinear eigenvalue problems in imaging. In Section 1.3.3 we extend those ideas from nonlinear image denoising to image segmentation. In Section 1.4 we describe the numerical realization of our approach using primal-dual convex optimization methods. In Section 1.5 we first illustrate the strengths and limitations of our multiscale segmentation method by studies on synthetic datasets with a certain focus on eigenfunctions and shapes. Moreover, we underline the potential and wide applicability by three different biomedical imaging applications. Its reliable performance is demonstrated on real fluorescence microscopy images that contain Circulation Tumor Cells, with various shapes and sizes, among white blood cells and debris in 1.5.4. Besides, we present results on electron microscopy images suffering from inhomogeneous backgrounds in 1.5.5 and interesting results on network-like shapes representing vascular systems in 1.5.6. We end with a conclusion and an outlook to future possible perspectives in Section 1.6.

1.2 Modeling Segmentation with Inverse Scale Spaces

In the following section we will shortly describe the model by Chan and Vese [24] for segmentation and the adaption of the ROF model for denoising using Bregman distances [50]. Afterwards we will introduce our novel Bregman-CV model for segmentation and show some advantages of our model.

1.2.1 Globally Convex Segmentation

The idea of the CV model has originally been derived from the more general model for image segmentation introduced by Mumford and Shah [47]. Here, one seeks for a solution of the variational energy

$$J^{\mathrm{MS}}(u, C) = \int_{\Omega} |f(x) - u(x)|^2 \mathrm{d}x + \alpha \cdot Per(C) + \beta \cdot \int_{\Omega \setminus C} |\nabla u(x)|^2 \mathrm{d}x \longrightarrow \min_{u,C}$$

where u is a differentiable function that is allowed to be discontinuous on C. C describes the union of the boundaries and thereby represents the contour defining the segmentation. Thus, u is a smooth approximation of the original image f and is composed of several regions Ω_i. Within each region Ω_i, u is smooth. If we restrict this model so that u is composed of

only two regions Ω_1 and Ω_2 we derive with $\beta \to \infty$ the CV model for segmentation

$$J^{\mathrm{CV}}(c_1, c_2, C) = \int_{\Omega_1} (f(x) - c_1)^2 \mathrm{d}x + \int_{\Omega_2} (f(x) - c_2)^2 \mathrm{d}x + \alpha \cdot Per(C) \longrightarrow \min_{C, c_1, c_2}.$$

Here, c_i is the intensity value of u within the corresponding region Ω_i. These values need to be determined together with the contour C. Note that in comparison to the original model we omit the area regularization term (cf. (1.1)).

The contour C can be indirectly represented via a level-set function Φ [51] with

$$C = \{x \in \Omega : \Phi(x) = 0\}$$

and $\Phi(x)$ being positive if and only if $x \in \Omega_1$. Together with the Heaviside function

$$H(\Phi(x)) = \begin{cases} 1 & \text{if } \Phi(x) \geq 0 \\ 0 & \text{if } \Phi(x) < 0 \end{cases}$$

and its regularized version H_ϵ this results in

$$J^{\mathrm{CV2}}(c_1, c_2, \Phi) = \int_\Omega (f(x) - c_1)^2 H_\epsilon(\Phi(x)) \mathrm{d}x + \int_\Omega (f(x) - c_2)^2 (1 - H_\epsilon(\Phi(x))) \mathrm{d}x + \alpha \cdot \int_\Omega |\nabla H_\epsilon(\Phi(x))| \mathrm{d}x \longrightarrow \min_{\Phi, c_1, c_2}.$$

The contour C evolves during minimization until it reaches a minimum which, in the ideal case, describes the object boundaries. Besides the original minimization strategy by gradient descent, several minimization methods to solve the CV model have been developed, see for example [5, 6, 37, 64]. One disadvantage of the model is its non-convexity which makes the solution depending on the used initialization. With a badly chosen initialization the minimization might get stuck in a local minimum that corresponds to a bad or meaningless segmentation.

For a better understanding of the relation between the nonlinear denoising model by Rudin, Osher and Fatemi [56] and the CV segmentation model we use the total variation defined as

$$TV(u) := \sup_{\substack{g \in C_0^\infty(\Omega; \mathbb{R}^2) \\ ||g||_\infty < 1}} \int_\Omega u \nabla \cdot g \, \mathrm{d}\mu \quad \text{with} \quad BV(\Omega) := \{u \in L^1(\Omega) | TV(u) < \infty\}. \tag{1.2}$$

The total variation of a characteristic $u(x) = \begin{cases} 1 & \text{if } x \in \Omega_1 \cup C \\ 0 & \text{if } x \in \Omega_2 \end{cases}$ corresponds to the contour length $|C|$ which can be shown by the co-area-formula. Therefore we can formulate the segmentation problem as

$$J^{CV3}(c_1, c_2, u) = \int_\Omega u \left((f(x) - c_1)^2 - (f(x) - c_2)^2 \right) \mathrm{d}x + \alpha \, TV(u) \longrightarrow \min_{\substack{u \in BV(\Omega), c_1, c_2 \\ u(x) \in \{0,1\}}}. \tag{1.3}$$

For fixed c_1, c_2 the solution of (1.3) corresponds to the solution of an ROF problem with a binary constraint [14]

$$\min_{\substack{u \in BV(\Omega) \\ u(x) \in \{0,1\}}} \frac{1}{2} ||u(x) - r(x)||_2^2 + \alpha \, TV(u) \tag{1.4}$$

with $r(x) = (f(x) - c_2)^2 - (f(x) - c_1)^2 + \frac{1}{2}$.

The regularization parameter α in the segmentation model (1.3) has the role of a scale parameter, meaning that α determines the scale of the objects that are segmented. The CV model describes a forward scale approach, thus a small parameter α corresponds to small scales that are segmented. An increased regularization parameter results in a solution where the smaller scales are not segmented but only larger ones. The meaning of scale is determined by the regularization functional, in this case the total variation. The total variation encodes a measure of the contour length as well as the height of piecewise constant areas. One disadvantage is that due to the 1-homogeneity of TV our method cannot distinguish between height and contour length. Thus, a small object with a bright intensity can have the same scale as a large object with a less bright intensity. For more details see section 1.5.

Convexification The CV segmentation model (1.3) as well as the binary ROF model (1.4) are both not convex. Even for fixed values of c_1 and c_2 both models are non-convex due to the binary constraint on u. As mentioned before, this might result in local instead of global minimum solutions. Approaches to overcome this difficulty and find global minima of the CV model are presented for example in [11, 12, 21, 35]. In [21] the authors showed that global minimizers of (1.3) for any given fixed $c_1, c_2 \in \mathbb{R}$ can be found by solving

$$J^{CV3}(c_1, c_2, v) = \int_\Omega v((f(x) - c_1)^2 - (f(x) - c_2)^2)dx + \alpha\ TV(v) \longrightarrow \min_{v \in BV(\Omega),\ v(x) \in [0,1]} \tag{1.5}$$

and defining $u(x) := v^*(x) \geq \mu$ for a.e. $\mu \in [0, 1]$.

Thus, the binary constraint can be relaxed and combined with a thresholding. Here, the variational model to solve is convex, though not strictly convex. One should bear in mind that the found solution is therefore not unique. Yet solutions of (1.5) are close to binary even if the constraint is relaxed. Thus for most choices of μ we derive the same solution which means that the choice of μ has only a very limited impact on our method. Therefore we do not see any disadvantages when choosing the global but not unique minimum $u(x) := v^*(x) \geq 0.5$. A fully convex formulation (including the constants c_1 and c_2) of problem (1.3) can be found in [12]. This method is computationally less efficient and currently we do not think that, in our method, the advantages of the full convexity outweigh the increased computational time.

1.2.2 Inverse Scale Space for TV-Denoising

Before introducing our new segmentation model in the following subsection we will first recall some properties of the well-known ROF model [56] and its extension by Bregman distances introduced in [50]. To denoise an image corrupted by additive Gaussian noise, [56] proposed to solve the nonlinear variational problem

$$\frac{1}{2}||u - f||_2^2 + \alpha\ TV(u) \longrightarrow \min_{u \in BV(\Omega)} \tag{1.6}$$

referred to as the ROF model. Similar to the CV model this generates a forward scale space flow regarding the scale parameter α. An increased parameter α leads again to a solution u

where fine scales are removed and vice versa. The total variation regularization functional is known to lead to a systematic contrast loss in the denoised image u [46]. To tackle this problem Osher et al. proposed in [50] an iterative contrast enhancement procedure based on Bregman distances. Instead of using the total variation regularization functional as before, information about the solution u that we gained from a prior solution of problem (1.6) is included. Therefore, problem (1.6) is replaced by a sequence of variational problems

$$u_{k+1} = \underset{u \in BV(\Omega)}{\operatorname{argmin}} \frac{1}{2}||u - f||_2^2 + \alpha\ D_{TV}^{p_k}(u, u_k). \tag{1.7}$$

The regularization $D_{TV}^{p_k}(u, u_k) := TV(u) - TV(u_k) - \langle p_k, u - u_k \rangle$ is the Bregman distance of u to the previous iterate u_k with respect to the total variation. $p_k \in \partial TV(u_k)$ is an element in the subdifferential of the total variation of the prior solution u_k. Although this subdifferential might be multivalued, the iterative regularization algorithm automatically selects a unique subgradient based on the optimality condition. For $k = 0$ we set $u_0 = p_0 = 0$. The iterative strategy of this model is as follows: We start with a large parameter α that results in an oversmoothed solution u_1 that consists of only large scales. In every iteration step finer scales are added back to the solution. Thus, the scale parameter that determines the range of the scales present in u is the iteration parameter k. In contrast to the forward approaches presented before, a small k corresponds to coarse scales and a large k to very fine scales. Therefore, the Bregman-ROF denoising approach is an inverse scale space approach. The authors showed that this strategy leads to enhanced contrast of the final solution $u_{k_{\max}}$ compared to the solution of (1.6). Hence, solving (1.7) instead of (1.6) with increasing α is not only an inverted way of detecting scales. This method rather allows for a detection of solutions which cannot be obtained by an adequate choice of α in the original ROF model.

1.2.3 Bregman-CV Segmentation Model

In the following section we will introduce our new inverse scale space approach for segmentation. It is based on the similarity of the ROF functional and the CV functional shown in (1.4). Similar to the Bregman-ROF denoising problem we replace the total variation regularization by an iterative regularization based on Bregman distances. Thus, the resulting novel segmentation model is given by

$$u_{k+1} = \underset{\substack{u \in BV(\Omega) \\ u(x) \in [0,1]}}{\operatorname{argmin}} \int_\Omega u \left((f - c_1)^2 - (f - c_2)^2 \right) + \alpha\ D_{TV}^{p_k}(u, u_k).$$

By inserting the definition of the Bregman distance $D_{TV}^{p_k}(u, u_k) := TV(u) - TV(u_k) - \langle p_k, u - u_k \rangle$ and ignoring the parts independent of u we derive the following model

$$u_{k+1} = \underset{u \in BV(\Omega)}{\operatorname{argmin}} \int_\Omega u \left((f - c_1)^2 - (f - c_2)^2 \right) + \chi_{[0,1]}(u) + \alpha\ \left(TV(u) - < u, p_k > \right) \tag{1.8}$$

with $p_k \in \partial TV(u_k)$, $p_0 = 0$ and $\chi_{[0,1]}(u) = 0$ if $u(x) \in [0, 1]$ and equal to infinity elsewhere. The range of the scales present in u_{k+1} is again determined by the iteration index k. This model

is an inverse scale space approach, thus a small k corresponds to a large scale segmentation and vice versa.

By definition of the Bregman distance it is $p_k \in \partial TV(u_k)$ where the subdifferential is multivalued. Therefore we need to determine a rule to choose p_k. One way is to derive an update strategy based on the optimality condition of (1.8) (cf. [50]):

$$0 = \left((f-c_1)^2 - (f-c_2)^2\right) + q_{k+1} + \alpha p_{k+1} - \alpha p_k \quad \text{(opt.cond.)}.$$

Here, q_{k+1} is an element in the subdifferential of the characteristic function $\chi_{[0,1]}(u_{k+1})$. The subdifferential of a characteristic function is a normal cone and is in our case given by

$$q_k(x) \in \begin{cases} (-\infty, 0] & \text{if} \quad u_k(x) = 0 \\ \{0\} & \text{if} \quad 0 < u_k(x) < 1 \\ [0, \infty) & \text{if} \quad u_k(x) = 1 \end{cases}.$$

Thus, we can choose $q_k = 0$ and neglect it from hereon. The update strategy for p_{k+1} is then given by

$$p_{k+1} = p_k - \frac{1}{\alpha}\left((f-c_1)^2 - (f-c_2)^2\right) = -\frac{k+1}{\alpha}\left((f-c_1)^2 - (f-c_2)^2\right). \tag{1.9}$$

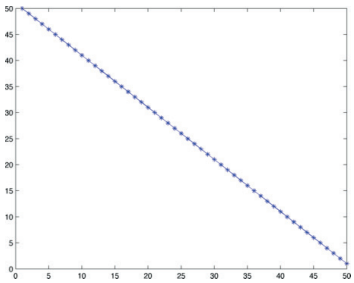
This update is independent of u_k, thus it is a pointwise constant update in every iteration.

1.2.4 Interpretation as an Adaptive Regularization Approach

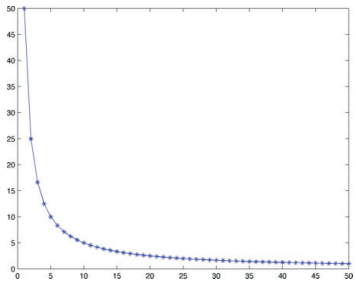
One important question is whether solutions of this model are in some sense improved compared to the solutions of the original CV model. We mentioned before that Bregman iterations lead to a contrast enhancement when applied to the ROF functional and that solving the CV model corresponds to solving a binary ROF. Yet, one should bear in mind that a contrast enhancement is meaningless in the case of a binary image since the contrast is already determined by the binary constraint. This is supported by the following observation: when inserting (1.9) into (1.8) we get

$$\begin{aligned} u_{k+1} &= \underset{\substack{u \in BV(\Omega) \\ u(x) \in [0,1]}}{\operatorname{argmin}} \int_\Omega u\left((f-c_1)^2 - (f-c_2)^2\right) + \alpha\left(TV(u) - < u, p_k >\right) \\ &= \underset{\substack{u \in BV(\Omega) \\ u(x) \in [0,1]}}{\operatorname{argmin}} \int_\Omega u\left((f-c_1)^2 - (f-c_2)^2\right) + \frac{\alpha}{(k+1)}\, TV(u) \end{aligned}$$

With this, it is straightforward to see that all solutions $u \in BV(\Omega)$ derived by the Bregman-CV model can also be found by the original CV model. Nevertheless, there are advantages of using the iterative update strategy. In Figure 1.2 (b) the resulting $\tilde{\alpha} = \frac{\alpha}{k+1}$ for $\alpha = 50$ and 50 Bregman iterations are shown. It is obvious that in the first iterations the decrease of the regularization parameter is much larger compared to later iterations. This is reasonable since first the large scales are reconstructed and in later Bregman iterations the finer scales are incorporated in the result. Making large steps with α in large scales and becoming finer is therefore a reasonable strategy. A large decrease in the later iterations would probably



(a) Linearly spaced α's from 50 to 1.



(b) "α's" resulting from 50 Bregman steps with $\alpha = 50$.

Figure 1.2: Comparison between linearly spaced regularization parameters and the automatically chosen parameters in the Bregman-CV model.

miss some scales in between while in the first iterations too small steps are not reasonable. With this strategy the problem of automatically choosing the regularization parameter is less severe. By choosing a large α and performing multiple Bregman iterations, a broad spectrum of scales is detected. Yet one should bear in mind that a strategy to automatically detect important scales is needed for a fully automated framework.

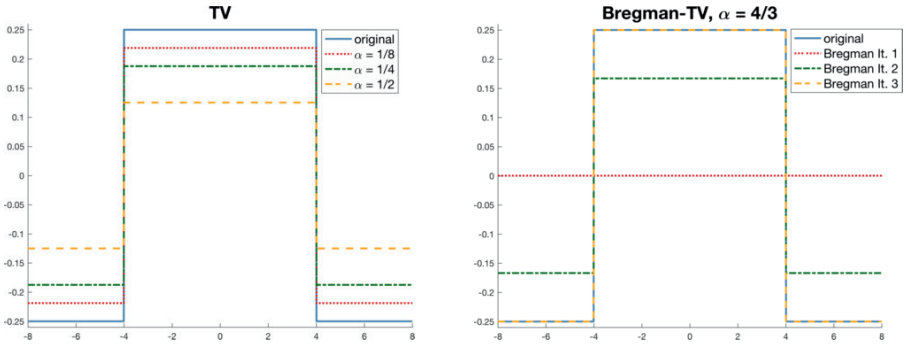
1.3 A Spectral Method for Multiscale Segmentation

The analysis of eigenvalues and spectral decomposition is a well-known theory in the field of linear signal and image processing (see e.g. [45] or [58] for a more recent overview). Since nonlinear regularization became popular in the last years, there is a growing interest in generalizing this theory to nonlinear operators. In [10] Benning et al. examined singular values for nonlinear, convex regularization functionals and in [32, 33] Gilboa transferred the idea of spectral decompositions to the nonlinear total variation functional and related operators. The general idea is to examine solutions of the nonlinear eigenvalue problem

$$\lambda u \in \partial J(u) \tag{1.10}$$

where ∂J denotes the subdifferential of a (one-homogeneous) convex functional usually representing regularization in inverse imaging problems. By transferring solutions of (1.10) to sparse peaks in a spectral domain, advanced filters enhancing or suppressing certain image components can easily be designed. This concept was first introduced for the total variation functional and later generalized to one-homogenous functions and analyzed in [15], [34] and [17].

In the case of a one dimensional signal the eigenfunction of the total variation corresponds to a single block with constant height, see Figure 1.3. When increasing the regularization parameter in the ROF model (1.6), i.e. when minimizing the total variation of this block signal, the block looses it height but the edges remain at the same position (cf. Figure 1.3 (a)). Therefore all signals can be seen as the original signal multiplied by a scalar and are



(a) Solution of the ROF model for different α.

(b) Solution of the inverse Bregman-ROF model for different Bregman iterations.

Figure 1.3: Solutions of the ROF and Bregman-ROF model in case the signal is a TV eigenfunction. Different regularization parameters α in the forward scale (left) and Bregman steps in the inverse scale space (right).

eigenfunction of the total variation. The same holds true for solutions of the Bregman-ROF model (1.7). The only difference is that now eigenfunctions of the TV functional (blocks) are reshaped with increasing number of Bregman iterations instead of removed (see Figure 1.3 (b)). This reflects the inverse scale approach of the Bregman-ROF model. Different to our novel approach, the Bregman updates in the Bregman-ROF model cannot be reformulated as an adaptive regularization parameter choice. As far as we know it is not entirely clear how the forward and inverse TV flow for denoising relate to each other.

1.3.1 Generalized Definition of the Total Variation

In the previous paragraph we mentioned that in the one dimensional case the eigenfunction of the total variation functional is a block with constant height. If we want to proceed to the two dimensional case the eigenfunctions are not as clearly defined as in the 1D case. Different than before we now have the freedom to choose the norm body that is used within the infinity norm in the TV definition (1.2). This choice determines the shape of the eigenfunctions. To reflect this dependence on the chosen norm in the TV definition we introduce a generalized version of the TV definition:

$$TV_\gamma(u) := \sup_{\substack{g \in C_C^1(\Omega;\mathbb{R}^d) \\ g(x)\cdot n < \gamma(n)\ \forall n \in \mathbb{R}^d}} -\int_\Omega u \nabla \cdot g \mathrm{d}x. \tag{1.11}$$

Here, $\gamma : \mathbb{R}^d \to \mathbb{R}$ is a convex, positively 1-homogeneous function such that $\gamma(x) > 0$ for $x \neq 0$. If u is a function in $\mathcal{W}^{1,1}(\Omega)$ the primal definition of equation (1.11) is given by

$$TV_\gamma(u) := \int_\Omega \gamma(\nabla u) \mathrm{d}x.$$

$\gamma(x)$	$\mathcal{F}_\gamma$	$\mathcal{W}_\gamma$
$\gamma = \|\cdot\|_1$		
$\gamma = \|\cdot\|_2$		
$\gamma = \|\cdot\|_\infty$		

Table 1.1: Examples for γ and the corresponding Frank diagrams $\mathcal{F}_\gamma$ and Wulff shapes $\mathcal{W}_\gamma$. The Wulff shape corresponds to the shape of the eigenfunctions of the TV_γ functional.

In both definitions the choice of γ determines the shape of the eigenfunctions. We refer to

$$\mathcal{F}_\gamma := \{z \in \mathbb{R}^d : \gamma(z) \leq 1\}$$

as the Frank diagram and the corresponding Wulff shape is defined as

$$\begin{aligned}\mathcal{W}_\gamma :=& \{z \in \mathbb{R}^d : z \cdot x \leq \gamma(x)\ \forall x \in \mathbb{R}^d\} \\ =& \{z \in \mathbb{R}^d : \gamma^*(x) := \sup_{x \in \mathbb{R}^d} \frac{z \cdot x}{\gamma(x)} \leq 1\}.\end{aligned}$$

Note that from definition (1.11) together with the definition of the Wulff shape we can conclude that $g(x) \in \mathcal{W}_\gamma$. Therefore, for every choice of γ, functions with the same shape as the Wulff shape are eigenfunctions of TV_γ (see [29] Theorem 4.1). The Frank diagram $\mathcal{F}_\gamma$ and the Wulff shape $\mathcal{W}_\gamma$ for the three most common choices of γ are presented in Table 1.1. For simplicity we will omit the subscript γ in TV_γ from hereon, but we will come back to it in the numerical results in section 1.5.

1.3.2 Spectral Analysis for Nonlinear Functionals

In the following section we will first review the basic ideas of spectral TV analysis before transferring those ideas to segmentation in the next section. The idea to decompose an image based on the basic TV elements was presented by Gilboa in [32, 33]. These basic TV elements, called eigenfunctions of the total variation functional, are all functions $u \in BV(\Omega)$ that solve the nonlinear eigenvalue problem

$$\lambda u \in \partial TV(u). \tag{1.12}$$

In the previous section we already presented some examples of TV eigenfunctions (see e.g. Figure 1.3 and Table 1.1). A more general description of these eigenfunctions in case γ is isotropic is given in [7]. Bellettini et al. showed that all indicator functions $1_C(x)$ of a convex and connected set C with finite perimeter $Per(C)$ which admit

$$\operatorname*{ess\,sup}_{p\in\partial C} \kappa(p) \leq \frac{Per(C)}{Area(C)} \tag{1.13}$$

where $Area(C)$ denotes the area of C and κ the curvature of $\partial C \in C^{1,1}$ are solutions of (1.12) and therefore eigenfunctions. Obtaining an analog condition for anisotropic, smooth and strictly convex choices of γ is straightforward but more challenging in the case of non-smooth or non-strictly convex choices of γ. Candidates of these shapes are presented in [8]. In [29] Esedoglu and Osher gave an example of a TV eigenfunction for $\gamma = \|\cdot\|_1$ that is not a Wulff shape (see section 1.5.3 for more details).

In order to detect eigenshapes at different scales in a given signal f, a scale space approach is needed. One way to define a scale space based on the total variation is given by the total variation flow [2, 3, 7, 16, 57]. The TV flow arises when minimizing the total variation with steepest descent method and is defined as

$$\begin{aligned} u_t(t,x) &= -p(t,x) \quad \text{for} \quad p(t,x) \in \partial TV(u(t,x)) \\ u(0,x) &= f(x) \end{aligned} \tag{1.14}$$

with Neumann boundary conditions. For $f(x) = 1_C(x)$, with C defined as above, the unique solution of (1.14) is $u(t,x) = (1-\lambda_C t)^+ 1_C(x)$ with $\lambda_C = \frac{Per(C)}{Area(C)}$ (cf. [7]). Hence, the time derivative $u_t(t,x)$ is given by the original signal multiplied with a scalar and u is an eigenfunction. To obtain a suitable framework to decompose or filter images based on these eigenfunction Gilboa proposed to define a spectral framework that transforms eigenfunctions to peaks in the spectral domain. If $u(t,x)$ is a solution of (1.14) the TV spectral transform and spectral response can be defined as

$$\phi(t,x) = u_{tt}(t,x)\cdot t \quad \text{and} \quad S(t) = ||\phi(t,x)||_{L^1(\Omega)}. \tag{1.15}$$

Note that there are alternative definitions of the spectral response presented in [15].

Another approach to construct an forward scale space is based on the variational ROF problem. Instead of solving (1.14) the ROF model

$$\min_{u\in BV(\Omega)} \frac{1}{2}||u(x) - f(x)||_2^2 + t\cdot TV(u) \tag{1.16}$$

is solved for different regularization parameters. Hence, t is the (artificial) time variable that determines the scale comparable to t in the TV-flow approach (1.14). One drawback of this approach is that there is no clear rule for the choice of different t's. The spectral transform and response function can be equivalently defined as in (1.15).

A third, but significantly different, scale space approach is an inverse scale approach. The inverse scale space flow is defined as

$$\begin{aligned} p_s(s,x) &= f(x) - u(s,x) \quad \text{for} \quad p(s,x) \in \partial TV(u(s,x)) \\ u(0,x) &= p(0,x) = 0. \end{aligned} \tag{1.17}$$

Thus, the flow is now defined on the dual variable $p \in \partial TV(u)$ and s is the time variable determining the scale. Note that in [18] it was shown that the iterative Bregman-L^2-TV model (1.7) can be associated with a discretization of (1.17) for $\frac{1}{\alpha} \to 0$. As (1.17) is an inverse approach the time variable t can be associated with $\frac{1}{s}$. In this case the spectral transform and response functions are defined as

$$\phi(s,x) = u_s(s,x) \quad \text{and} \quad S(s) = ||\phi(s,x)||_{L^1(\Omega)}, \tag{1.18}$$

where $u(s,x)$ is the solution of (1.17). Note that for small s $\phi(s,x)$ now measures changes in the coarse scales. See [15] for more details.

With all three approaches we are able to transform a signal to the spectral domain and detect different scales based on TV eigenfunctions. If we assume that $\phi(t,x)$ is integrable over time, the original signal f can be reconstruct via

$$f(x) = \int_0^\infty \phi(t,x)dt + \bar{f}$$

where $\bar{f}$ is the average of f. Filters can be defined with $\phi_H(t,x) = H(t)\phi(t,x)$ via

$$f_H(x) = \int_0^\infty \phi_H(t,x)dt + H(\infty)\bar{f}.$$

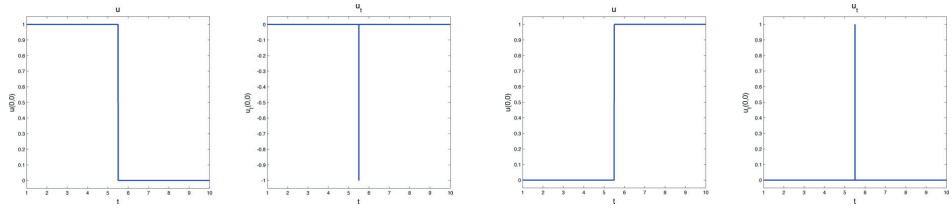
1.3.3 Spectral Response of Multiscale Segmentation

In section 1.2.3 we presented two variational models to detect segmentations of a given image f at different scales. To decompose the segmentation into different scales and detect important scales or clusters of scales in the segmentation we want to transfer the idea of spectral analysis based on the total variation (cf. sec. 1.3.2) to segmentation. Therefore we need to find a suitable transformation of the segmentation u to the spectral domain and vice versa. Note that our goal is not to reconstruct the original signal f or filtered versions of f but the reconstructed function should be a segmentation itself. To do so we make use of the idea that eigenfunctions of the TV functional should be transformed to peaks in the spectral domain. In the following we will derive this spectral transform function for a forward scale space and an inverse scale space approach. To represent the forward scale space we associate the regularization parameter α in the convex version of the original model by Chan and Vese (1.5) with the artificial time variable t. That means, we solve

$$\int_\Omega u((f(x)-c_1)^2 - (f(x)-c_2)^2)dx + t \cdot TV(u) \longrightarrow \min_{u\in BV(\Omega),\ u(x)\in[0,1]} \tag{1.19}$$

for different t. This is comparable to the variational approach in (1.16). So far, we could not find a forward scale space representation that can be associated with a flow on u comparable to the TV-flow. One difficulty is that the optimality condition of this model, i.e. $0 = (f-c_1)^2 - (f-c_2)^2 + \alpha p$ with $p \in TV(u)$, has no direct dependence on u. The inverse scale space representation is based on the Bregman-CV model we introduced in (1.8). The optimality condition in each step of the iterative Bregman-CV strategy is given as

$$0 = \left((f-c_1)^2 - (f-c_2)^2\right) + \alpha\,(p_k - p_{k-1}) \quad \text{with } p_k \in TV(u_k)\ \forall\ k.$$



(a) Evolution of u in (1.19). (b) Evolution of u in (1.20).

Figure 1.4: Illustration of the evolution of $u \in \{0,1\}$ over time at a fixed point when f is a TV eigenfunction shown for the forward (a) and inverse scale space approach (b). Note due to the binary constraint on u the evolution of u is already a step function, i.e. $u(x)$ is either part of the segmentation (u(x)=1) or not (u(x)=0).

The resulting equation

$$\frac{p_k - p_{k-1}}{\frac{1}{\alpha}} = (f - c_2)^2 - (f - c_1)^2$$

can be interpreted as a discretization with stepsize $\frac{1}{\alpha}$ of

$$\begin{aligned} p_s(s,x) &= (f(x) - c_2)^2 - (f(x) - c_1)^2 \quad \text{with} \quad p(s,x) \in \partial TV(u(s,x)) \\ p(0,x) &= 0. \end{aligned} \tag{1.20}$$

Again, s is the time variable that is inverse to the time variable t in (1.19). We refer to this flow as the inverse scale space segmentation flow. Note that within this flow description there is no direct dependence on u but u is only indirectly given by $p \in \partial TV(u)$. Yet, when looking for solutions of this flow at different times s, we solve the Bregman-CV model for multiple Bregman iterations and there the corresponding u is available.

To transform the eigenfunctions to peaks in the spectral domain we define the spectral transform function as follows

$$\phi(t,x) = \begin{cases} -u_t(t,x) & \text{(forward case)} \\ u_t(t,x) & \text{(inverse case).} \end{cases} \tag{1.21}$$

This definition is motivated by Figure 1.4 where the evolution of TV eigenfunctions over time in (1.19) and (1.20) is illustrated. We can see, that the first derivative (in a distributional sense) is a peak at that time point where the eigenfunction vanishes or appears respectively. For the spectral response function we use the definition

$$S(t) = ||\phi(t,x)||_{L^1(\Omega)} \tag{1.22}$$

introduced by Gilboa [32, 33]. Here, the influence of certain scales to the segmentation is encoded. Using this function ϕ and assuming integrability over time we can get the following backtransformation

$$f^{\text{seg}}(x) = \int_0^\infty \phi(t,x)dt,$$

where $f^{\text{seg}}(x) := (f-c_2)^2 - (f-c_1)^2 + \frac{1}{2} < \frac{1}{2}$ is the results of the simple clustering problem $\min_{u(x)\in\{0,1\}} \int_\Omega u\left((f(x)-c_1)^2 - (f(x)-c_2)^2\right) \mathrm{d}x$. By choosing a function H with $H(t) \in \{0,1\}$ certain scales of interest can be filtered via

$$f_H^{\text{seg}}(x) = \int_0^\infty \phi_H(t,x)dt \quad \text{with} \quad \phi_H(t,x) = H(t)\phi(t,x). \tag{1.23}$$

With this framework we can easily segment an image and simultaneously detect important scales in this segmentation instead of using the spectral TV analysis as in Section 1.3.2 and segmenting separately. Moreover, using the filtering approach we can easily construct segmentations of only certain scales by filtering (cf. (1.23)). Examples are shown in section 1.5.

1.4 Numerical Methods

Our novel approach introduced in the two previous sections consists of two parts. First we have to solve the Bregman-CV model (1.8) introduced in section 1.2.3. Afterwards we can analyze the detected scales using the spectral response function (1.22) introduced in section 1.3.3. Therefore an efficient solution of the Bregman-CV model is required. In the following section we will give a very brief introduction into primal-dual optimization schemes and then show how this can be used to solve (1.8). We close the section by a speed comparison between a MATLAB® and a parallelized C/MEX implementation of our code.

1.4.1 Primal-Dual Optimization Methods

To solve nonlinear problems of the form

$$\min_{x\in X} \; F(Kx) + G(x) \tag{1.24}$$

with F and G being proper, convex, lower-semicontinuous functions, primal-dual optimization methods became very popular in the last years. Instead of minimizing the primal problem (1.24) they make use of the primal-dual formulation of this problem given by

$$\min_{x\in X} \max_{y\in Y} \langle Kx, y\rangle + G(x) - F^*(y) \tag{1.25}$$

with $G : X \to [0,+\infty]$ and $F^* : Y \to [0,+\infty]$ being the convex-conjugate of F. By updating in every iteration step a primal and a dual variable these methods are able to avoid some difficulties that arise when working on the purely primal or dual problem. One example is the minimization of variational methods with TV regularization. If the gradient is zero, the TV functional is not differentiable which leads to problems in purely primal minimization schemes like gradient descent. Some examples for primal-dual minimization algorithms are the PDHG algorithm [66], a generalization of PDHG by Esser et al. [30], the Split Bregman algorithm by Goldstein and Osher [36], Bregman iterative algorithms [63], the second-order CGM algorithm [22] and inexact Uzawa methods [65].

In [20] Chambolle and Pock proposed an algorithm which can be seen as a generalization of PDHG as well. This algorithm was originally proposed by Pock et al. in [53] to minimize

a convex relaxation of the Mumford-Shah functional. The efficient first-order primal-dual algorithm to minimize general problems of the form (1.25) is presented in Algorithm 1. $(I + \sigma\partial F^*)^{-1}$ and $(I + \tau\partial G)^{-1}$ are the resolvent operators of F^* and G respectively which are defined through

$$y = (I + \tau\partial G)^{-1}(x) = \operatorname*{argmin}_y \left\{ \frac{\|y - x\|^2}{2\tau} + G(y) \right\}.$$

Algorithm 1 First-order primal-dual algorithm by Chambolle and Pock [20]

Parameters: $\tau, \sigma > 0$, $\theta \in [0, 1]$
Initialization: $n = 0,\ x^0 \in X,\ y^0 \in Y, \bar{x}^0 = x^0$

Iteration:

for $(n \geq 0)$ **do**

1. $y^{n+1} = (I + \sigma\partial F^*)^{-1}(y^n + \sigma K\bar{x}^n)$.
2. $x^{n+1} = (I + \tau\partial G)^{-1}(x^n - \tau K^* y^{n+1})$.
3. $\bar{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n)$.
4. Set $n = n + 1$.

end for

return x^n

1.4.2 Numerical Realization Bregman-CV

In order to use Algorithm 1 to solve the constraint problem (1.8) we reformulate the problem into

$$\int_\Omega u\left((f - c_1)^2 - (f - c_2)^2\right) + \mathrm{id}_{[0,1]}(u) + \alpha\ (TV(u) - < u, p_k >) \longrightarrow \min_{u \in BV(\Omega)} \tag{1.26}$$

with $p_k \in \partial TV(u_k)$ and $p_0 = 0$. $\mathrm{id}_{[0,1]}(u)$ is the indicator function of the interval $[0, 1]$ defined as 0 if $u \in [0, 1]$ and ∞ otherwise. To derive an minimization strategy based on Algorithm 1 we set $x = u$, $K(u) = \nabla u$ and

$$F(u) = \|u\|_1 \text{ and } G(u) = \mathrm{id}_{[0,1]}(u) + \int_\Omega u\left((f - c_1)^2 - (f - c_2)^2 - \alpha p_k\right).$$

The convex-conjugate of $F(u) = \|u\|_1$ is given by $F^*(g) = \delta_P(g)$ with $P = \{g : \|g\|_\infty \leq 1\}$ and

$$\delta_P(g) = \begin{cases} 0 & \text{if } g \in P \\ \infty & \text{if } g \notin P \end{cases}. \tag{1.27}$$

Together with (1.25) we derive the primal-dual variant of (1.26):

$$\langle \nabla u, g\rangle + \mathrm{id}_{[0,1]}(u) + \int_\Omega u\left[(f-c_1)^2 - (f-c_2)^2 - \alpha p_k\right] - \alpha\delta_P(g) \longrightarrow \min_u \max_g . \tag{1.28}$$

The resolvent operators for G and F^* are defined through

$$\begin{aligned} u = (I+\tau\partial G)^{-1}(\tilde{u}) &= \mathrm{Proj}_{[0,1]}\left[\tilde{u} - \tau\left((f-c_1)^2 - (f-c_2)^2 - \alpha p_k\right)\right] \\ &= \max\left(0, \min\left(1, \tilde{u} - \tau\left((f-c_1)^2 - (f-c_2)^2 - \alpha p_k\right)\right)\right) \end{aligned} \tag{1.29}$$

and

$$g = (I+\sigma\partial F^*)^{-1}(\tilde{g}) = \mathrm{Proj}_{\{\{g:\|g\|_\infty\le 1\}\}}\left(\tilde{g_{ij}}\right). \tag{1.30}$$

Note that the L^∞ norm $\|g\|_\infty$ is in the discrete setting defined as $\|g\|_\infty = \max_{i,j}\{\gamma^*(g_{i,j})\}$ where the choice of γ determines the shape of the eigenfunctions of the TV functional. For $\gamma = \|\cdot\|_{\ell^p}$ with $p = 1$ the unit ball defined by

$$\{(x,y)\in\Omega|\max\{|x|,|y|\}\le 1\}$$

is an TV eigenfunction, for $p = 2$ the unit ball defined by

$$\left\{(x,y)\in\Omega|\sqrt{|x|^2+|y|^2}\le 1\right\}$$

and for $p = \infty$ the unit ball defined by

$$\{(x,y)\in\Omega|(|x|+|y|)\le 1\}.$$

However these are not the only eigenfunctions. With (1.29) and (1.30), we derive the primal-dual algorithm presented in Algorithm 2 to minimize (1.8). Note that we are not updating the constants c_1 and c_2, but start with a good estimate and leave it fixed. To a certain extend the varying regularization parameter can compensate for an error in those constants. Some examples are presented in Section 1.5.

We implemented two versions of the code, the first one is written in MATLAB® and the second one is written in C and called via MATLAB® using the MEX interface. For speed comparisons we've tested both code versions (always with 1 Bregman iteration) on two different machines. First, a MacBook Pro with an Intel Core i5 processor with two cores of 2.6 GHz and 8GB of memory (solid graphs in Figure 1.5) and the second machine was a Dell Linux workstation with in Intel Xeon processor with 16 cores of 2.6GHz and 32GB of memory (dashed graphs in Figure 1.5). Moreover, we ran the C/MEX code with and without parallelization using openMP. The results of our speed test can be seen in Figure 1.5. We see that especially for large images the computational time is significantly decreased when using the C implementation (yellow and green graphs) compared to the pure MATLAB® code (blue graphs). All C versions behave very similar for small images but deviate for larger images. Without parallelization the C code is slightly faster on the Mac (solid yellow graph). When activating openMP and parallelizing parts of the code the computational time could be further improved for both computers. The difference is much more significant when using the computer with more cores. For the Mac we see a speed-up of about 30 percent (solid green graph) for large images while on the Linux computer (dashed green graph) we could achieve a speed-up of about 65 percent compared to the C implementation without parallelization.

Algorithm 2 First-order primal-dual algorithm to solve (1.8).

Parameters: data f, reg. param. $\alpha \geq 0$, $\tau, \sigma > 0$, $\theta \in [0,1]$, $maxIts \in \mathbb{N}$, $maxBregIts \in \mathbb{N}$
Initialization: $l = 1$, $u_1^0 = 0$, $p_0 := 0, \bar{u}^0 = u^0$

Iteration:

while $(k < maxBregIts)$ **do**

1. Set n=0.

   while $(n < maxIts)$ **do**

   a) $g^{n+1} = \text{Proj}_{\{\{g:\|g\|_\infty \leq 1\}\}}(g^n + \sigma \nabla \bar{u}^n)$.

   b) $u^{n+1} = \max\left(0, \min\left(1, u^n + \tau \nabla \cdot g^{n+1} - \frac{\tau}{\alpha}\left((f-c_1)^2 - (f-c_2)^2 - \alpha p_k\right)\right)\right)$.

   c) $\bar{u}^{n+1} = u^{n+1} + \theta(u^{n+1} - u^n)$.

   d) Set $n = n + 1$.

   end while

2. Update $p_{k+1} = p_k + \frac{1}{\alpha}\left((f-c_2)^2 - (f-c_1)^2\right)$.
3. Set $u_{k+1}^0 = u_k^n$.
4. Set $k = k + 1$.

end while

return u_k^0

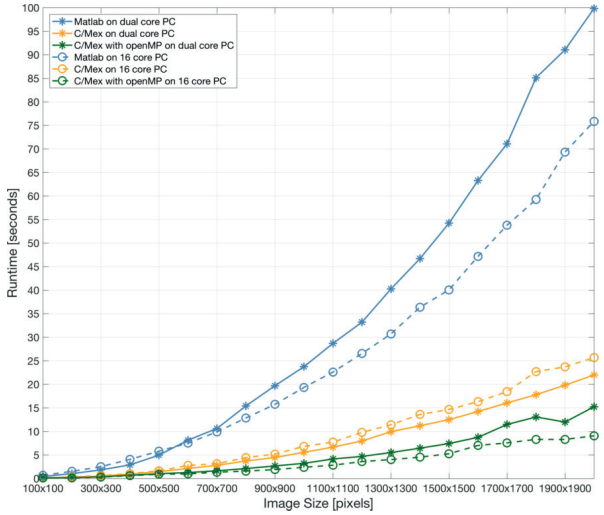


Figure 1.5: *Comparison of runtimes for MATLAB® and C/MEX code.* The code was tested on 2 different computers: a Mac with an Intel Core i5 processor with two cores of 2.6 GHz and 8GB of memory (solid graphs) and a Linux computer with an Intel Xeon processor with 16 cores of 2.6GHz and 32 GB of memory (dashed graphs). On both computers we have tested a full MATLAB® implementation (blue), a C/MEX implementation without parallelization (yellow) and a C/MEX implementation with parallelization using openMP (green).

1.5 Experimental Results

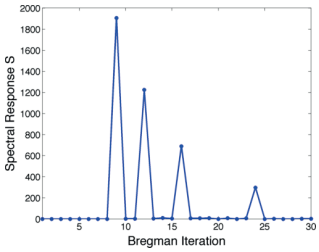
In this section we illustrate the main properties, advantages and limitations of our novel multiscale segmentation framework via several synthetic experiments as well as real experiments from biological cell imaging and retina imaging. The main framework that we apply and study throughout this section consists of Algorithm 2, performing the inverse scale space method Bregman-CV in (1.8), together with an evaluation of the spectral response S in (1.22).

In the first set of segmentation experiments we focus on simple objects (e.g. discs) and investigate how underlying object intensities and sizes can automatically be detected as multiple scales. In the second part we concentrate on the stability of our method under uncertainty like noise. In the third part of this section we generalize our study to other simple shapes (e.g. blocks or diamonds), explore the connection to underlying eigenfunctions of the nonlinear TV_γ functional and the potential for shape detection respectively shape optimization. Finally, we apply our framework to real images obtained from the CellSearch system and other automatic scanning microscopes used for the enumeration of Circulating Tumor Cells [1] offering new automatic procedures for tumor cell identification, quantification and classification. Moreover, we show potential applications of our algorithm in the field of electron microscopy of biological samples and retina imaging.

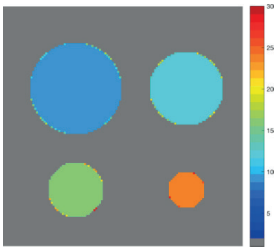
1.5.1 Multiscale Segmentation: Size versus Intensity

In this first subsection we start with the investigation of basic segmentation tasks. We try to find round objects (discs) with varying scale (e.g. differences in size or intensity) in front of homogeneous backgrounds to verify the basic properties of our model. Since discs are the unit balls of the ℓ^2 norm we chose $\gamma = \gamma^* = \|\cdot\|_2$ in the TV norm (1.11).

(a) 1st Bregman iteration. (b) 9th Bregman iteration. (c) 12th Bregman iteration. (d) 16th Bregman iteration. (e) 24th Bregman iteration.



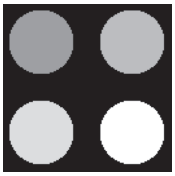
(f) Spectral response S.



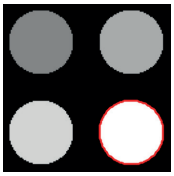
(g) Color-coded segmentation of multiple scales.

Figure 1.6: *Detection of size scales*. Automatic segmentation of discs of different sizes under fixed intensity (height). Segmentation is visualized via red contour. (a)-(e) Segmentation after different number of Bregman Iterations. (f) Associated spectral response function S. The number of Bregman iterations is plotted on the x-axis. Every peak in S corresponds to one object that appears in the segmentation (compare to (a) - (e)). The decreasing size of the peaks reflects the decreasing disc size.

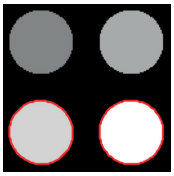
In the first example in Figure 1.6 (a) we can automatically identify four discs with different diameters (sizes) via the Bregman-CV inverse scale space algorithm 2. As visualized in Figure 1.6 (b)-(e) the method automatically picks out the discs one by one while iterating (w.r.t $t \approx \alpha_k$) through the scale space. At the end, the full foreground-background segmentation is available. In addition to this, the spectral response (1.22) of our iterates regarding t, nicely reflects the sparse occurrence of new contours, see Figure 1.6 (f). Note that in this example the intensity of peaks decreases due to the different size of each disc. The height of S at each time point reflects the number of pixels that occurred (or vanished) in the segmentation in this step. In Figure 1.6 (g) we plotted the function $\int_t \phi(x,t)\cdot t\mathrm{d}t$ in color coding. To obtain an easily interpretable visualization we use a gray background. Since the function $\phi(x,t)$ for given x is equal to 1 exactly at the time point t_x where x occurs in the segmentation and 0 elsewhere, $\int_t \phi(x,t)\cdot t\mathrm{d}t$ corresponds for every point x to t_x. We can see that apart from the



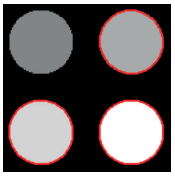
(a) 1st Bregman iteration.



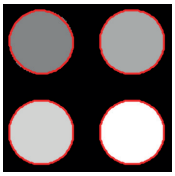
(b) 3rd Bregman iteration.



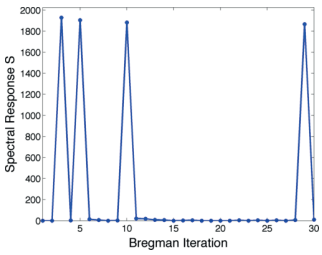
(c) 5th Bregman iteration.



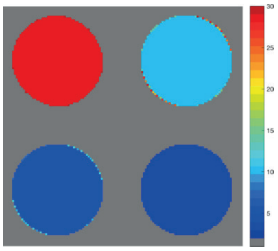
(d) 10th Bregman iteration.



(e) 29th Bregman iteration.



(f) Spectral response S.



(g) Color-coded segmentation of multiple scales.

Figure 1.7: *Detection of intensity scales.* Automatic segmentation of discs of different intensities (heights) under fixed size. Segmentation is visualized via red contour. (a)-(e) Segmentation after different number of Bregman Iterations. (f) Associated spectral response function S. The number of Bregman iterations is plotted on the x-axis. Every peak in S corresponds to one object that appears in the segmentation (compare to (a) - (e)).

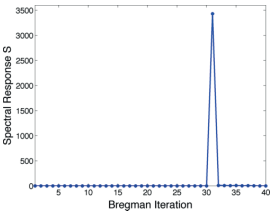
four time steps where the four discs appear nearly nothing is appearing in between, reflecting the sparse behavior of our methods. Only very few pixels at the object boundaries occur in later iterations which is caused by discretization inaccuracies. In this way the multiscale approach of our method can be seen since each color reflects one segmented scale.

The results of the same experiment but now with intensity scales are shown in the next figure. The four discs in Figure 1.7 (a) all have the same size but vary in their intensity from 0.55 to 1. Note that in some sense this contradicts the assumption of the CV model that the data roughly consists of two different intensity levels c_1 and c_2. Nevertheless our multiscale approach is able to detect all scales reliably. Like in all our examples we estimated the constants beforehand as $c_1 = \text{mean}\{f(x)|f(x) < \frac{1}{2} \cdot \max f(x)\}$ and $c_2 = \text{mean}\{f(x)|f(x) \geq \frac{1}{2} \cdot \max f(x)\}$ and leave them fixed throughout the whole iteration process. Figure 1.7 (b)-(e) visualizes the result of the four Bregman iterations in which the different discs appear. The color-coded result of multiples scales ($\int_t \phi(x,t) \cdot t \mathrm{d}t$) is shown in Figure 1.7 (g). Again, only a few pixels occur at time steps apart from the four significant ones. It is remarkable that the first three scales appear shortly after each other while the last scale appears strikingly later. This can be explained by the way the regularization parameter is adapted automatically through the Bregman distance (cf. Figure 1.2). In later iterations the change in α_k becomes smaller so that the time until a certain object appears is larger

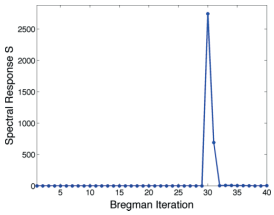
(a) Data f with intensity of 0.68 (large disc).

(b) Data f with intensity of 0.69 (large disc).

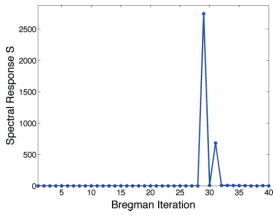
(c) Data f with intensity of 0.70 (large disc).



(d) Spectral response for (a).



(e) Spectral response for (b).



(f) Spectral response for (c).

Figure 1.8: *Scale ambiguity w.r.t. size and intensity.* (a) - (c) Three different datasets that differ only in the intensity of the larger circle (0.68/0.69/0.7). The associated spectral responses S are shown below. (a) together with (d) shows that out method is not able to reliably distinguish size and intensity scales but only a slight decrease of the intensity leads to two separate peaks in the spectral response function ((b) together with (e) and (c) together with (f)).

compared to the previous iterate. Although this might lead to late appearance of smaller scales we can thereby guarantee to not miss scale differences in the finer scales. The late appearance of the last disc is also reflected in the spectral response function S in Figure 1.7 (f). Again, the response function has a sparse representation and the last peak occurs delayed. Apart from small variations caused by discretization artifacts the peaks now have all the same height reflecting their similar size.

Limitations: In the previous two examples we showed how we can reliably detect different size scales and different intensity scales with our method. One limitation of our approach is the simultaneous detection of size and intensity scales in an image. Figure 1.8 (a)-(c) shows three different data sets, each with a small disc with intensity 1 and a larger discs with intensity $0.68 - 0.7$ respectively. For all three examples we computed the spectral response with fixed parameters $\alpha = 200$ and 40 Bregman iterations shown in Figure 1.8 (d)-(f). We can see that in Figure 1.8 (d) the spectral method was not able to detect the difference between a large object with a lower intensity and a smaller object with a high intensity. Both objects appear at exactly the same point in time. Figure 1.8 (e) and (f) shows that already a small increase of the lower intensity circumvents the problem and both objects are represented

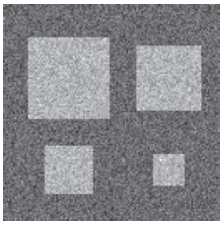
as individual peaks in the spectral response. The height of each peaks reveals which object occurs. Yet we hope to solve this issue in future versions; solution ideas are presented in the summary and conclusion section 1.6.

1.5.2 Robustness against Noise

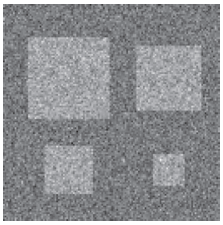
In this second subsection we investigate the robustness of our methods with respect to uncertainties like noise. Figure 1.9 (a)-(d) shows the same underlying image but with different noise levels. The underlying signal is a binary signal showing four squares with different size. We added Gaussian noise with mean 0 and standard deviation $\sigma = 0.25, 0.5, 0.75$ and 1 respectively. Since squares are the unit balls of the ℓ^∞ norm we chose $\gamma = \|\cdot\|_1$ and therefore $\gamma^* = \|\cdot\|_\infty$. The results of our inverse Bregman-CV model are plotted in terms of the spectral response function S and the color-coded segmentation of all scales $\int_t \phi(x,t)\cdot t\,\mathrm{d}t$. Note that we chose the same parameter setting for all four examples ($\alpha = 100$ and 30 Bregman iterations). In Figure 1.9 (e)-(h) we can see that the spectral response function has for all noise levels a sparse representation indicating very clearly the four scales in the underlying signal. It is also remarkable that the peaks are roughly at the same points in time, shifting only slightly to the right for higher noise levels. This is even more obvious in the color-coded representation Figure 1.9 (i)-(l). The color bar next to each image shows that the color coding is consistent over all four examples. Thus, the same colors (see for example Figure 1.9 (i) and (j)) reflect that the objects occur at the exact same iteration unaffected of the changed noise level. Even in case of high noise (σ = 0.75) only the second and the fourth largest object appear one iteration later although the smaller objects are now affected by an incomplete segmentation. In case of very high noise ($\sigma = 1$) we cannot perfectly segment the objects, especially the smaller ones "get lost" in the noise. Nevertheless the spectral response function is still close to sparse and for all four objects we reliably segment most of the object. These results show that our method is very robust to noise and circumvents to some extend the problem of choosing a good regularization parameter α. The parameter setting that we used yielded a reliable extraction of all scales in the underlying signal independent of the noise present in our data.

1.5.3 Segmentation regarding Shapes of Eigenfunctions

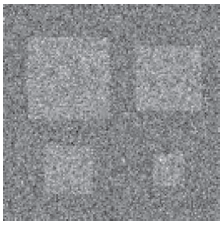
This subsection is about the choice of the γ-function and how this choice is reflected in our spectral analysis framework. As mentioned in Section 1.3.1 by varying γ in the generalized definition of the total variation (1.11) the shape of the eigenfunctions is changed. In the following two examples we will first show different examples of eigenfunctions that lead to a sparse representation of the spectral response function. Here we differentiate between eigenfunctions that are Wulff shapes of the chosen γ function and those that are not. In addition, we investigate what happens if the choice of γ does not match the shapes present in the image.



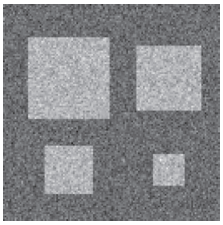
(a) Noise level σ = 0.25.



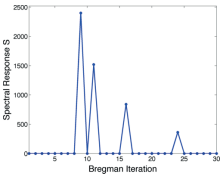
(b) Noise level σ = 0.5.



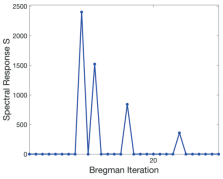
(c) Noise level σ = 0.75.



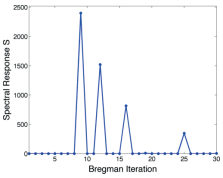
(d) Noise level σ = 1.



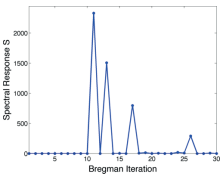
(e) Spectral response S.



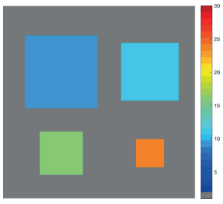
(f) Spectral response S.



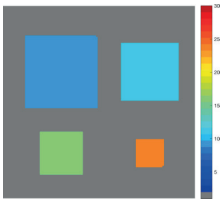
(g) Spectral response S.



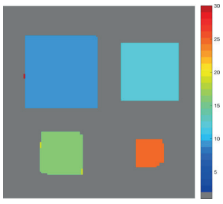
(h) Spectral response S.



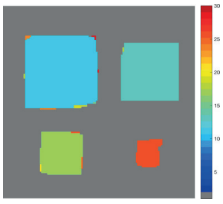
(i) Color-coded segmentation of multiple scales.



(j) Color-coded segmentation of multiple scales.

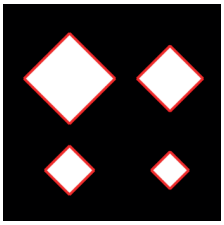


(k) Color-coded segmentation of multiple scales.

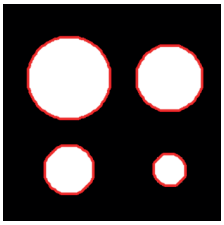


(l) Color-coded segmentation of multiple scales.

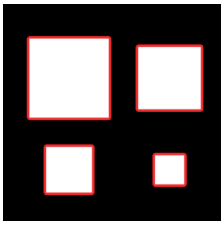
Figure 1.9: *Robustness against noise*. (a) - (d) shows the same underlying binary signal but with different levels of noise added. Exactly the same model parameters are used to process all 4 datasets. The associated spectral response functions are shown in (e) - (h) and the color-coded segmentation results are shown in (i) - (l). We can see that our method is very robust against noise, we only see a slight shift to the right in the spectral response function for higher noise levels. For high noise levels some artifacts at the objects boundary occur but we reliably find all 4 objects independent of the noise level.



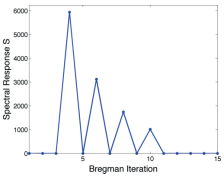
(a) Data f with segmentation shown in red.



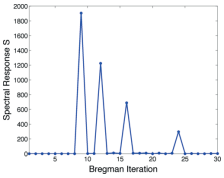
(b) Data f with segmentation shown in red.



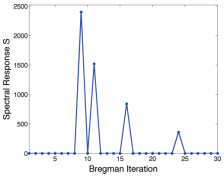
(c) Data f with segmentation shown in red.



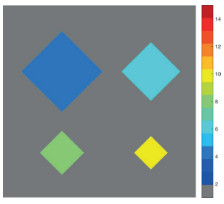
(d) Spectral response S.



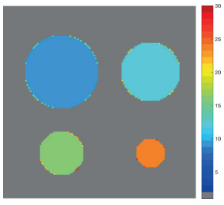
(e) Spectral response S.



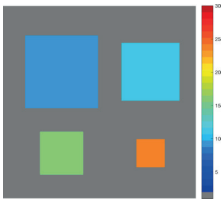
(f) Spectral response S.



(g) Color-coded segmentation of multiple scales.



(h) Color-coded segmentation of multiple scales.

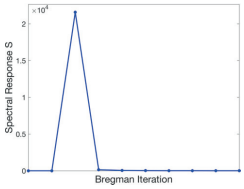


(i) Color-coded segmentation of multiple scales.

Figure 1.10: *Varying norm bodies and eigenshapes.* Automatic detection of different signals composed of eigenshapes for different norm bodies. Column one shows the results for $\gamma^* = \|\cdot\|_1$, column two for $\gamma^* = \|\cdot\|_2$ and column three for $\gamma^* = \|\cdot\|_\infty$. (a) - (c) Segmentation results visualized via red contour after 30 Bregman Iterations. Associated spectral response function S is shown in (d) - (f) and the color-coded segmentation in (g) - (i). With the correct choice of γ all eigenshapes can be detected in a single step.

In Figure 1.10 (a)-(c) each images shows a composite of eigenfunction for three different ℓ^p-norm choices for γ (namely $p = \infty$, 2 and 1). In each of the images, the input data is shown together with the final segmentation result in red. The associated spectral responses are shown below (Figure 1.10 (d)-(f)). We can see that as long as the choice of γ and the given dataset match (in the sense that the components of f have the same shape as the unit-ball of γ^* (cf. Table 1.1)) we get a sparse response function. In Figure 1.10 (g)-(i) the color-coded segmentation of all scales is shown. We can see that for rectangular shapes ((g) and (i)) no discretization artifacts at the boundary occur. Moreover, we can again use the same

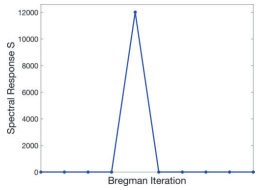
(a) TV eigenfunction for $\gamma = \|\cdot\|_2$.



(b) Spectral response S.

(c) Color-coded segmentation.

(d) TV eigenfunction for $\gamma = \|\cdot\|_1$.



(e) Spectral response S.

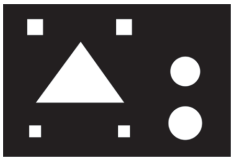
(f) Color-coded segmentation.

Figure 1.11: *Reconstruction of eigenfunctions that are not Wulff shapes.* (a) and (d) show two TV eigenfunctions that are not Wulff shapes for $\gamma = \|\cdot\|_2$ and $\gamma = \|\cdot\|_1$ respectively. The associated spectral responses are shown in (b) and (e) and the color-coded segmentation results in (c) and (f). It is obvious that eigenfunctions that are not Wulff shapes do have a sparse spectral response representation and appear in one step in the corresponding segmentation.

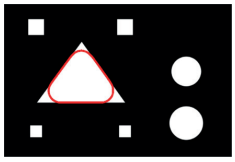
parameter setting for all three data sets. The similar color coding in (h) and (i) indicates that the objects occur roughly at the same time point independent of their shape and γ. The data set in the first column is larger than the other ones thus the objects occur earlier and we need less iterations.

In Figure 1.11 (a) and (d) we present two eigenfunctions which are not Wulff shapes. The round object in (a) fulfills condition (1.13) and is therefore a TV eigenfunction for $\gamma = \|\cdot\|_2$. The corresponding spectral response function in (b) and the color-coded segmentation in (c) show that not only eigenfunctions that are Wulff shapes appear in one step but also more general eigenfunctions. Another example (now for $\gamma = \|\cdot\|_1$) is shown in Figure 1.11 (d). Esedoglu and Osher showed in [29] that a rectangle whose contours are parallel to the x- and y-axis respectively is an eigenfunction of the anisotropic TV functional. This is reflected in the spectral response function Figure 1.11 (e) and the color-coded segmentation in 1.11 (f) which show that the rectangle appears in one step although it is not a Wulff shape of the 1-norm. For natural images, where objects are often not Wulff shapes, this allows objects with various shapes to appear in one step. Some examples can be seen in the subsections 1.5.4 - 1.5.6.

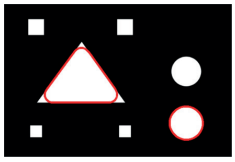
In Figure 1.12 we investigated how our multiscale approach behaves in case of shape mixtures and shapes that do not match the prior shape assumption made in γ. The given data (Figure 1.12) consists of different size scales and different sizes with only two objects



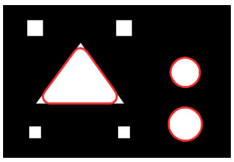
(a) Given Data.



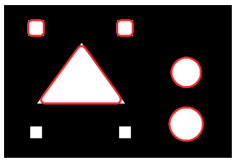
(b) Bregman iteration 10.



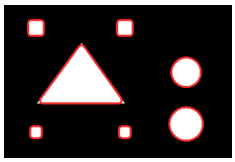
(c) Bregman iteration 16.



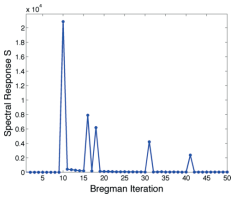
(d) Bregman iteration 18.



(e) Bregman iteration 31.



(f) Bregman iteration 41.



(g) Spectral response S.

(h) Color-coded segmentation of multiple scales.

Figure 1.12: *Reconstruction of mixed shapes.* Automatic detection of a signal consisting of different size scales and shapes. The binary signal is shown in (a). In (b) - (f) the segmentation results of the Bregman iterations where a significant peak in the spectral response S in (g) appeared are shown. The color-coded segmentation is shown in (h). Here the behavior of non-eigenfunction over time can be seen. They do not appear in one step but reshape and evolve into the shape of the original signal.

matching the prior assumption of round objects ($\gamma = \|\cdot\|_2$). Figure 1.12 (b)-(f) shows all time steps where a new scale occurs in the segmentation (shown in red). We can see that the round objects appear in one step while all other ones appear with rounded edges and reshape over time. This can be seen more clearly in the color-coded segmentation of all steps shown in Figure 1.12 (h). Especially for the large triangle in the middle we can nicely see how non-eigenfunctions are reshaped over time and how the segmentation is very round at the beginning and then propagates into the edges. Nevertheless we cannot get clear edges, neither for the triangle nor for the squares. The spectral transform function is nevertheless close to sparse but due to this reshaping behavior there is a small signal also between the clear peaks. It might be suitable to use the ℓ_0-norm of S as a measure in order to evaluate if a certain signal matches the prior shape assumption made. However for this idea the influence of discretization artifacts as shown in some examples before, would have to be clarified in further detail.

1.5.4 Fluorescence microscopy images containing Circulating Tumor Cells

In this subsection we demonstrate how our multiscale segmentation approach can be applied to the analysis of Circulating Tumor Cells (CTCs). Before presenting some experimental data sets we will first give a short introduction on CTCs, associated research questions and challenges for diagnosis and treatment of cancer patients.

Circulating Tumor Cells are cells that dissociate from a primary tumor and invade the blood stream. In recent literature it was shown that the number of CTCs present in the blood is associated with the survival chance of a patient and can be used to guide therapy of cancer patients. The elimination of CTCs indicates effective therapy whereas increase or failure to eliminate indicates a futile therapy. A challenge in the identification of CTCs is that they are very rare and therefore difficult to detect among other cells in the pool of blood cells. The gold standard for CTC enumeration is the CellSearch system and in prospective multicenter studies a threshold of 5 CTCs / 7.5 ml of blood was used to separate patients with metastatic disease into those with favorable and unfavorable prognosis [26, 28]. Currently all known CTC analysis tools are based on subjective morphological criteria and objects can be classified differently by different operators which results in different interpretation of the status of the patients. The low cell counts and subsequent statistical analyses further increases the chances of mistakes. The problem of subjectivity in the CTC analysis is currently addressed by the development of an open-source toolbox to automatically detect and classify CTCs in datasets from various machines and institutes. This project is part of a European consortium called CANCER-ID that aims at validating blood-based biomarkers and is funded by the European Union. One key component of this toolbox is the development of a (nearly) parameter-free segmentation algorithm which performs reliably and efficiently on multiple tumor cell datasets (see Figure 1.14). Moreover, in recent studies [25] it was hypothesized that not only the definition used to classify objects as CTC predicted clinical outcome but also those objects that did not match the criteria of an intact cell. A schematic overview of different cell classes and their relative frequencies is shown in Figure 1.13. Here, green is a fluorescence marker indicating the presence of cytokeratin expressed in cancer cells but not on white blood cells and magenta is a fluorescence marker indicating the presence of DNA thus showing the nucleus of the cell. When comparing the green marker among the different CTC classes we observe that the signal clusters strongly vary in size (area) and therefore motivate the usage of our multiscale segmentation approach.

In Figure 1.14 (a), (d), (g) and (j) experimental data sets are shown. From the original image sets, consisting of three different fluorescent color channels, we extracted the tumor cell marker (green) and used those images as input for our multiscale segmentation approach. Although the difficulties vary between all images (inhomogeneous background, noise, cell clusters, mixture of size and intensity scales), we can process all images with our multiscale segmentation approach with the *exact same parameters*. This is essential for the development of a user-friendly (parameter-free) toolbox for CTC analysis. Note that the dim spots in image (d) are not cells but only pores of a filter used to collect cells (bright spots) and therefore it is not desired to segment them. The resulting spectral response functions for

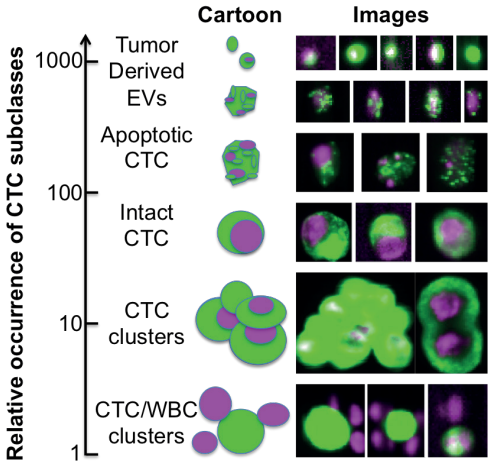
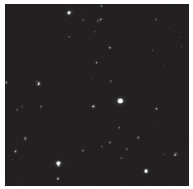
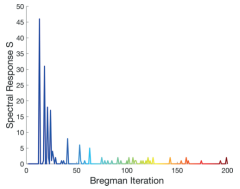


Figure 1.13: *Schematic overview of different CTC classes.* For each subclass a cartoon image of a typical cell is shown. Next to it on the right different examples for every class are shown. All images consist of two overlaid color channels where green is the tumor cell marker and magenta is the nucleus marker. The relative frequencies of each subclass are shown on the left. By looking only at the green signal it is obvious that the size scale varies between different subclasses.

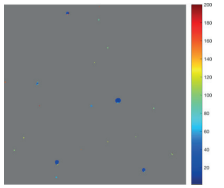
all four images are shown in Figure 1.14 (b), (e), (h) and (k) with a color coding corresponding to the coding used in the segmentation results in (c), (f), (i) and (l). The color coding of the response function shows that all objects which appear later in our segmentation and therefore belong to finer scales have a yellow to reddish color in the color-coded segmentation. The very large and intact cells are blue (with some small artifacts at the boundary) and smaller cells (or large fragments) are shown in light blue to green. We can nicely observe that the object colors cover the whole color scale range. For images that are more complex (e.g. (d) and (j)) also the spectral response function is more complex but the color-coded segmentation shows that nearly every object appears in one step and thereby has a clearly defined scale that we use as a feature in our classification approach. Here, we profit from the fact that not only Wulff shapes (perfectly circular objects) but also other eigenshapes appear in one step in our segmentation. Hence, this segmentation approach not only provides all contours automatically without any parameter adaptations but simultaneously also a simple classification of cells based on their size (scale resp. color/appearance time). This analysis can be applied to all color channels separately and be used together with more features in a subsequent automatic classification approach. The constants c_1 and c_2 can again simply be estimated a-priori from the data by a simple thresholding and averaging approach and are fixed throughout the iterative process.



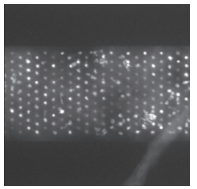
(a) Fluorescence image from the CellSearch system.



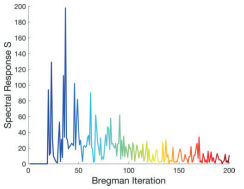
(b) Color-coded spectral response S.



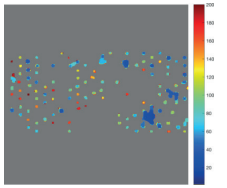
(c) Color-coded segmentation of multiple scales.



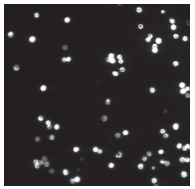
(d) Fluorescence image from a filter system.



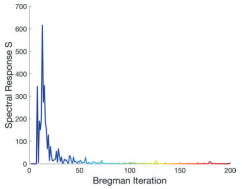
(e) Color-coded spectral response S.



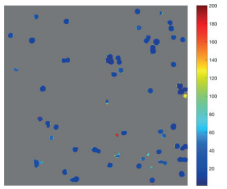
(f) Color-coded segmentation of multiple scales.



(g) Fluorescence image from the CellSearch system with more clustered cells.

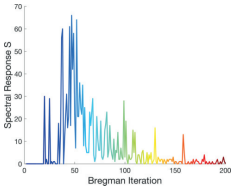


(h) Color-coded spectral response S.

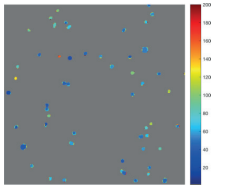


(i) Color-coded segmentation of multiple scales.

(j) Fluorescence image from the ARIOL system.

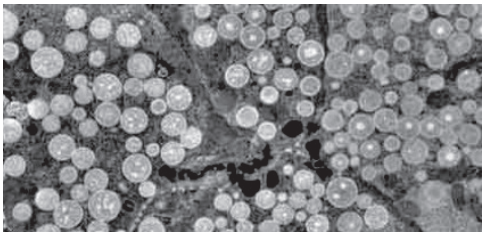


(k) Color-coded spectral response S.

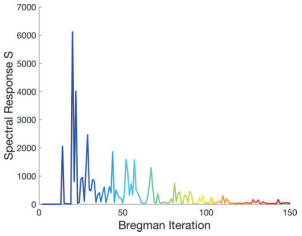


(l) Color-coded segmentation of multiple scales.

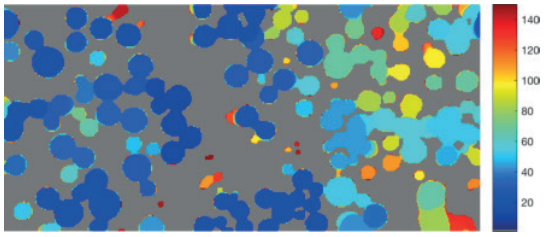
Figure 1.14: *Experimental cell data.* Automatic detection of (a) different sparse cell sizes (scales) with nearly constant background, (d) cells in front of an inhomogeneous background, (g) clustered cells with nearly constant background and (j) different intensity and size scales. All results are obtained with the exact same parameters. In column one single channel images obtained from different fluorescence microscopes are shown. The corresponding spectral responses are shown in the second column and the color-coded segmentation in column three. It can be seen that more complex images (inhomogeneous background or mixtures of scales) lead to more complexity in the spectral response function (response becomes less sparse).



(a) Electron microscopy data of acinar cells in the parotid gland provided by [55].



(b) Color-coded spectral response S.



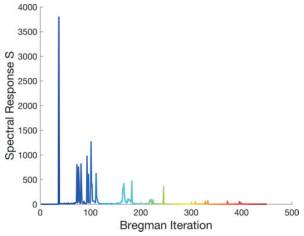
(c) Color-coded segmentation of multiple scales.

Figure 1.15: *Segmentation of cells in front of an inhomogeneous background. (a) shows the input image obtained by electron microscopy, (b) the corresponding spectral response function and (c) the color-coded segmentation. This result shows that our method can detect inhomogeneities in the object we want to segment while inhomogeneities in the background (up to a certain degree) do not affect the result of our segmentation method.*

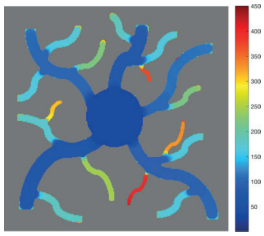
1.5.5 Electron microscopy cell images with nonuniform background

In this subsection we show that our multiscale segmentation approach can also be used for real-world images where the background is less homogenous than in the cell images before. In Figure 1.15 (a) we show an electron microcopy dataset of acinar cells provided by [55]. For a more obvious visual interpretation we reverted the contrast from the original EM dataset (bright background with dark cells) to a dark, but not uniform, background with brighter cells. The intensity of the cells is also inhomogeneous and varies from very bright cells (left) to cells with an intensity closer to the background (right). Figure 1.15 (b) shows the color-coded spectral response function and (c) the color-coded segmentation of the cells. We see that the inhomogeneous intensity of the cells is reflected in the scales found by our method. The inhomogeneity of the background is not influencing the segmentation result and the constants c_1 and c_2 can still be estimated beforehand from the data without further adaption. One problem that occurs is the ambiguity of size and intensity scales which can be observed when comparing small but bright cells on the left (visualized in green) with larger cell clusters on the right (also visualized in green). Currently our method is not able to differentiate their scales. A useful extension of our method would be a combination of our multiscale approach with a watershed type algorithm that is able to split cell clusters into individual cells of different scales.

(a) Given Data.



(b) Color-coded spectral response *S*.



(c) Color-coded segmentation.

Figure 1.16: *Reconstruction of different scales in network-like structures.* The input data is shown in (a) with the corresponding spectral response function (b) and the color-coded segmentation (c). In this example the usability of our method in case of more complex, network-like structures is demonstrated. Interestingly, the spectral response function shows distinct sparse clusters where every cluster corresponds to "arms" or "branches" of a specific diameter. Although they are not shaped as an eigenfunction they appear (nearly) in one step and are represented by one specific scale.

1.5.6 Segmentation and clustering of biological network structures

In many biological and medical applications network-like structures occur as part of vascular systems. Examples are blood-vessels in retinal images, anisotropic structures in brain or myocardium imaging or at the microscopic level even cells consisting of a round core with several meandering arms of different diameters. Although these networks have a much more complex shape than the examples before, our multiscale method is very useful for a segmentation or clustering of these networks. An example of a round object with "arms" of different diameters is shown in Figure 1.16 (a). In (b) the corresponding spectral response function is shown and the color-coded segmentation in (c). We can see that the spectral response function has sparse clusters where each cluster corresponds to branches of a certain diameter. Although these "arms" are not eigenfunctions, they do appear in one step and therefore have a distinct scale if they have constant diameter. A similar behavior can be seen in Figure 1.17. As input for our multiscale method we used a manual blood vessel segmentation (b) of a retinal image (a) from the STARE dataset [38]. The resulting spectral response is shown in Figure 1.17 (c) and the color-coded segmentation in (d). We see that the different times of appearance of the vessels (indicated by the different colors) lead to a clustering of the underlying network based on the diameter of each branch. With this approach we can get segmentations of the network that contain only branches of certain diameter scales or directly get an estimate of how many different vessel sizes are present and how often they occur. In order to apply our segmentation approach directly on raw retinal images, instead of using it as a postprocessing step, a decomposition method for retinal images similar to Zosso [67] could be combined with our multiscale approach.

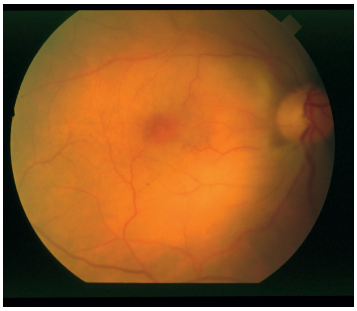
1.6 Summary and Conclusion

In this paper, we have studied a novel multiscale segmentation method which is a combination of an inverse scale space variational approach and a nonlinear spectral analysis. Using our approach we can automatically detect objects of different scales where scale can refer to intensity or size. These scales can be evaluated using a spectral response function and we can easily decompose the segmentation with respect to those scales using the spectral transform function. Thus we have extended the useful spectral analysis, that was introduced for TV denoising applications before, to a segmentation approach. This novel approach can to some extent circumvent the problem of choosing a suitable regularization parameter. The iterative procedure using Bregman distances can be interpreted as an automatic regularization parameter choice method and is very robust with respect to noise or errors in the estimated c_1 and c_2 values in the CV model. We have shown that our method works very reliable even for applications with more than two intensity scales although this actually contradicts the models binary assumption. The framework can easily be adapted to different prior shape assumptions (Wulff shapes). With regard to real experimental datasets such an assumption is not limiting the usefulness of the method, due to its ability to even detect compositions of eigenfunctions not belonging to the restricted class of Wulff shapes. For the numerical implementation we used a primal-dual optimization scheme that, especially with a C/MEX version, lead to a very efficient framework, even for large datasets. We illustrated the strengths and limitations of our method on synthetic datasets with a certain focus on eigenfunctions, shapes as well as robustness against noise and background artifacts. Three different biomedical imaging applications with different shapes and challenges emphasize the potential and wide applicability of our approach. The code is also part of an open-source toolbox called ACCEPT. With this software users are able to automatically detect and classify CTCs from fluorescence images, but can also extract important information related to the level of expression of certain therapy targets on the CTC (see the application and results described in section 1.5.4).

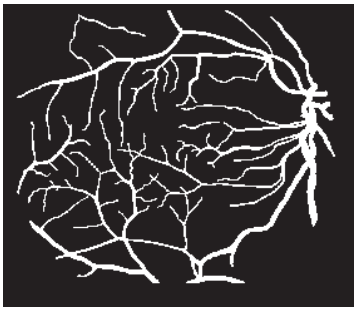
One limitation that we currently have in our model is an ambiguity in scale. It is known for variational methods with an L^2 type data fidelity term combined with total variation regularization that these methods are not able to distinguish between intensity and size scales. In practice it seems not very likely that two objects with different size and different intensity fall onto exactly the same peak in the spectral response. Nevertheless to reliably interpret the spectral response function it is advantageous to have only one dominating scale in the data that is analyzed. Another limitation is the flexibility with regard to different shapes. Currently only one shape assumption at the moment is possible.

Outlook. There are four main questions that we want to address in our future research. To allow a clear interpretation of the spectral response function even in the case of size and intensity scales we want to adapt the dataterm in the way that it can differentiate both scales. For denoising this can be done using an L^1 type data fidelity and we want to generalize this to our segmentation model (which is not obvious in terms of gradient flows). For the experimental cell datasets we currently analyze every fluorescent channel separately. In a future work we would like to investigate if a combined segmentation in all channels

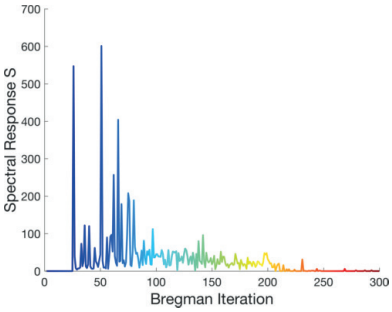
can lead to an improved segmentation and a richer response function. Moreover, we are interested in how far we can use the spectral response function to investigate if a shape assumption matches the data. One idea is to use the l_0 "norm" of the spectral response function as a measure for shape optimization. A further development of this idea could lead to an approach where we could learn an optimal γ function and thereby an optimal shape of eigenfunctions for a specific dataset. Another very interesting consideration is to transfer our method from the currently local setup to a nonlocal setup. A first foundation for this was recently laid with a work on non-local TV spectral theory by Aujol et al. in [4]. More mathematical spectral theory for nonlocal TV, its usefulness for solving and improving complex segmentation tasks and a direct comparison with spectral clustering and graph cuts will be of high interest for the community and our future research.



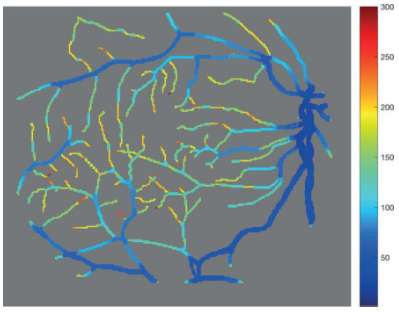
(a) Retinal image of the STARE dataset [38].



(b) Manually segmented blood vessel network [38].



(c) Color-coded spectral response S.



(d) Color-coded clustering of network in (b).

Figure 1.17: *Clustering of networks.* (a) A retinal image of a human eye taken from the STARE dataset [38] with a manual segmentation of the blood vessel network in (b). We applied our multiscale segmentation method to the binary image in (b) and thereby obtain the spectral response function (c) and a clustering of the network based on the diameter of the network branches (d). We see that large blood vessels are represented by blue colors, medium vessels by light blue to green colors and the very fine branches by yellow to red colors. Although the representation of (c) is not as sparse as in the previous example we can get a meaningful clustering of the graph based on the resulting colors in (d).

Bibliography

[1] W. J. Allard, J. Matera, M. C. Miller, M. I. Repollet, M. C. Connelly, C. G. Rao, A. G. Tibbe, J. W. Uhr, and L. W. Terstappen. Tumor Cells Circulate in the Peripheral Blood of All Major Carcinomas but not in Healthy Subjects or Patients with Nonmalignant Diseases. *Clinical Cancer Research*, 10(20):6897–6904, 2004.

[2] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing Total Variation Flow. *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics*, 331(11):867–872, 2000.

[3] F. Andreu, V. Caselles, J. I. Díaz, and J. M. Mazón. Some Qualitative Properties for the Total Variation Flow. *Journal of Functional Analysis*, 188(2):516–547, feb 2002.

[4] J.-F. Aujol, G. Gilboa, and N. Papadakis. Fundamentals of Non-Local Total Variation Spectral Theory. In *Scale Space and Variational Methods in Computer Vision*, pages 66–77. Springer, 2015.

[5] N. Badshah and K. Chen. Multigrid Method for the Chan-Vese Model in Variational Segmentation. *Communications in Computational Physics*, 4(2):294–316, 2008.

[6] E. Bae and X.-C. Tai. Efficient Global Minimization for the Multiphase Chan-Vese Model of Image Segmentation. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 28–41, Berlin, Heidelberg, 2009. Springer.

[7] G. Bellettini, V. Caselles, and M. Novaga. The Total Variation Flow in RN. *Journal of Differential Equations*, 184(2):475–525, 2002.

[8] G. Bellettini, M. Novaga, and M. Paolini. Characterization of Facet Breaking for Nonsmooth Mean Curvature Flow in the Convex Case. *Interfaces and Free Boundaries*, 3(4):415–446, 2001.

[9] M. Benning, C. Brune, M. Burger, and J. Müller. Higher-Order TV Methods - Enhancement via Bregman Iteration. *Journal of Scientific Computing*, 54(2-3):269–310, 2013.

[10] M. Benning and M. Burger. Ground States and Singular Vectors of Convex Variational Regularization Methods. *Methods and Applications of Analysis*, 20(4):295–334, 2013.

[11] X. Bresson, S. Esedoglu, P. Vandergheynst, J.-P. Thiran, and S. J. Osher. Fast Global Minimization of the Active Contour/Snake Model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, 2007.

[12] E. S. Brown, T. F. Chan, and X. Bresson. Completely Convex Formulation of the Chan-Vese Image Segmentation Model. *International Journal of Computer Vision*, 98(1):103–121, 2012.

[13] C. Brune, A. Sawatzky, and M. Burger. Primal and Dual Bregman Methods with Application to Optical Nanoscopy. *International Journal of Computer Vision*, 92(2):211–229, 2011.

[14] M. Burger, Y. Dong, and M. Hintermüller. Exact Relaxation for Classes of Minimization Problems with Binary Constraints. *arXiv:1210.7507*, oct 2012.

[15] M. Burger, L. Eckardt, G. Gilboa, and M. Moeller. Spectral Representations of One-Homogeneous Functionals. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 16–27. Springer, 2015.

[16] M. Burger, K. Frick, S. J. Osher, and O. Scherzer. Inverse Total Variation Flow. *Multiscale Modeling & Simulation*, 6(2):366–395, 2007.

[17] M. Burger, G. Gilboa, M. Moeller, L. Eckardt, and D. Cremers. Spectral Decompositions Using One-Homogeneous Functionals. *SIAM Journal on Imaging Sciences*, 9(3):1374–1408, 2016.

[18] M. Burger, S. J. Osher, J. Xu, and G. Gilboa. Nonlinear Inverse Scale Space Methods for Image Restoration. In *Variational, Geometric, and Level Set Methods in Computer Vision*, pages 25–36, 2005.

[19] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.

[20] A. Chambolle and T. Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[21] T. F. Chan, S. Esedoglu, and M. Nikolova. Algorithms for Finding Global Minimizers of Image Segmentation and Denoising Models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, jan 2006.

[22] T. F. Chan, G. H. Golub, and P. Mulet. A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration. *SIAM Journal on Scientific Computing*, 20(6):1964–1977, 1999.

[23] T. F. Chan, B. Y. Sandberg, and L. A. Vese. Active Contours without Edges for Vector-Valued Images. *Journal of Visual Communication and Image Representation*, 11(2):130–141, 2000.

[24] T. F. Chan and L. A. Vese. Active Contours Without Edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[25] F. A. Coumans, C. J. Doggen, G. Attard, J. S. de Bono, and L. W. Terstappen. All Circulating EpCAM+ CK+ CD45- Objects Predict Overall Survival in Castration-Resistant Prostate Cancer. *Annals of Oncology*, 21(9):1851 – 1857, 2010.

[26] M. Cristofanilli, G. T. Budd, M. J. Ellis, A. Stopeck, J. Matera, M. C. Miller, J. M. Reuben, G. V. Doyle, W. J. Allard, L. W. Terstappen, and D. F. Hayes. Circulating Tumor Cells, Disease Progression, and Survival in Metastatic Breast Cancer. *New England Journal of Medicine*, 351(8):781–791, 2004.

[27] E. B. Dam and M. Nielsen. Non-Linear Diffusion for Interactive Multi-Scale Watershed Segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 216–225. Springer, 2000.

[28] J. S. de Bono, H. I. Scher, R. B. Montgomery, C. Parker, M. C. Miller, H. Tissing, G. V. Doyle, L. W. Terstappen, K. J. Pienta, and D. Raghavan. Circulating Tumor Cells Predict Survival Benefit From Treatment in Metastatic Castration-Resistant Prostate Cancer. *Clinical Cancer Research*, 14(19):6302–6309, 2008.

[29] S. Esedoglu and S. J. Osher. Decomposition of Images by the Anisotropic Rudin-Osher-Fatemi Model. *Communications on Pure and Applied Mathematics*, 57(12):1609–1626, 2004.

[30] E. Esser, X. Zhang, and T. F. Chan. A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science. *SIAM Journal on Imaging Sciences*, 3(4):1015–1046, 2010.

[31] L. Florack and A. Kuijper. Topological Structure of Scale-Space Images. *Journal of Mathematical Imaging and Vision*, 12(1):65–79, 2000.

[32] G. Gilboa. A Spectral Approach to Total Variation. In A. Kuijper, K. Bredies, T. Pock, and H. Bischof, editors, *Scale Space and Variational Methods in Computer Vision*, pages 36–47. Springer, 2013.

[33] G. Gilboa. A Total Variation Spectral Framework for Scale and Texture Analysis. *SIAM Journal on Imaging Sciences*, 7(4):1937–1961, jan 2014.

[34] G. Gilboa, M. Moeller, and M. Burger. Nonlinear Spectral Analysis via One-Homogeneous Functionals: Overview and Future Prospects. *Journal of Mathematical Imaging and Vision*, 56(2):300–319, 2016.

[35] T. Goldstein, X. Bresson, and S. J. Osher. Geometric Applications of the Split Bregman Method: Segmentation and Surface Reconstruction. *Journal of Scientific Computing*, 45(1-3):272–293, 2010.

[36] T. Goldstein and S. J. Osher. The Split Bregman Method for L1-Regularized Problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[37] L. He and S. J. Osher. Solving the Chan-Vese Model by a Multiphase Level Set Algorithm Based on the Topological Derivative. In *Scale Space and Variational Methods in Computer Vision.*, pages 777–778. Springer, 2007.

[38] A. Hoover, V. Kouznetsova, and M. Goldbaum. Locating Blood Vessels in Retinal Images by Piecewise Threshold Probing of a Matched Filter Response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000.

[39] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active Contour Models. *International Journal of Computer Vision*, 1(4):321–331, 1988.

[40] S. Kichenassamy, A. Kumar, P. Olver, A. Tannenbaum, and A. J. Yezzi. Gradient Flows and Geometric Active Contour Models. In *Proceedings of IEEE International Conference on Computer Vision*, pages 810–815. IEEE, 1995.

[41] J. J. Koenderink. The Structure of Images. *Biological Cybernetics*, 50(5):363–370, 1984.

[42] M. M. Letteboer, O. F. Olsen, E. B. Dam, P. W. Willems, M. A. Viergever, and W. J. Niessen. Segmentation of Tumors in Magnetic Resonance Brain Images Using an Interactive Multiscale Watershed Algorithm. *Academic Radiology*, 11(10):1125–1138, 2004.

[43] L. M. Lifshitz and S. M. Pizer. A Multiresolution Hierarchical Approach to Image Segmentation Based on Intensity Extrema. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(6):529–540, 1990.

[44] T. Lindeberg. Feature Detection with Automatic Scale Selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[45] S. L. Marple Jr. *Digital Spectral Analysis: with Applications*. Prentice-Hall, Inc., 1987.

[46] Y. Meyer. *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, volume 22. American Mathematical Soc., 2001.

[47] D. Mumford and J. Shah. Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[48] W. J. Niessen, K. L. Vincken, J. A. Weickert, B. M. ter Haar Romeny, and M. A. Viergever. Multiscale Segmentation of Three - Dimensional MR Brain Images. *International Journal of Computer Vision*, 31(2-3):185–202, 1999.

[49] W. J. Niessen, K. L. Vincken, J. A. Weickert, and M. A. Viergever. Nonlinear Multiscale Representations for Image Segmentation. *Computer Vision and Image Understanding*, 66(2):233–245, 1997.

[50] S. J. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An Iterative Regularization Method for Total Variation-Based Image Restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[51] S. J. Osher and J. A. Sethian. Fronts Propagating with Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

[52] K. Papafitsoros and K. Bredies. A Study of the One Dimensional Total Generalised Variation Regularisation Problem. *Inverse Problems and Imaging*, 9(2):511–550, 2015.

[53] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An Algorithm for Minimizing the Mumford-Shah Functional. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1133–1140. IEEE, 2009.

[54] C. Pöschl and O. Scherzer. Exact Solutions of One-Dimensional Total Generalized Variation. *Communications in Mathematical Sciences*, 13(1):171–202, 2015.

[55] D. Riedel. Electron Microscopy Image.

[56] L. I. Rudin, S. J. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[57] G. Steidl, J. A. Weickert, T. Brox, P. Mrázek, and M. Welk. On the Equivalence of Soft Wavelet Shrinkage, Total Variation Diffusion, Total Variation Regularization, and SIDEs. *SIAM Journal on Numerical Analysis*, 42(2):686–713, 2004.

[58] P. Stoica and R. Moses. *Spectral Analysis of Signals*. Prentice Hall, 2005.

[59] M. Tabb and N. Ahuja. Multiscale Image Segmentation by Integrated Edge and Region Detection, 1997.

[60] L. A. Vese and T. F. Chan. A Multiphase Level Set Framework for Image Segmentation Using the Mumford and Shah Model. *International Journal of Computer Vision*, 50(3):271–293, 2002.

[61] K. L. Vincken, A. S. Koster, and M. A. Viergever. Probabilistic Multiscale Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):109–120, 1997.

[62] A. Witkin. Scale-Space Filtering: A New Approach to Multi-Scale Description. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 9, pages 150–153. IEEE, 1984.

[63] W. Yin, S. J. Osher, D. Goldfarb, and J. Darbon. Bregman Iterative Algorithms for L1 - Minimization with Applications to Compressed Sensing. *SIAM Journal on Imaging Sciences*, 1(1):143–168, 2008.

[64] N. E. Zehiry, S. Xu, P. Sahoo, and A. Elmaghraby. Graph cut Optimization for the Mumford-Shah Model. In *The Seventh IASTED International Conference on Visualization, Imaging and Image Processing*, pages 182–187, 2007.

[65] X. Zhang, M. Burger, and S. J. Osher. A Unified Primal-Dual Algorithm Framework Based on Bregman Iteration. *Journal of Scientific Computing*, 46(1):20–46, 2011.

[66] M. Zhu and T. F. Chan. An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration. Technical Report 1, CAM 08-34, UCLA, 2008.

[67] D. Zosso, J. An, J. Stevick, N. Takaki, M. Weiss, L. S. Slaughter, H. H. Cao, P. S. Weiss, and A. L. Bertozzi. Image Segmentation with Dynamic Artifacts Detection and Bias Correction. *Inverse Problems and Imaging*, 11(3):577–600, 2017.

Chapter 2

Combining Contrast Invariant L1 Data Fidelities with Nonlinear Spectral Image Decomposition

Leonie L. Zeune, Stephan A. van Gils, Leon W.M.M. Terstappen and Christoph Brune

International Conference on Scale Space and Variational Methods in Computer Vision, Springer (2017): 80-93

Abstract

This paper focuses on multi-scale approaches for variational methods and corresponding gradient flows. Recently, for convex regularization functionals such as total variation, new theory and algorithms for nonlinear eigenvalue problems via nonlinear spectral decompositions have been developed. Those methods open new directions for advanced image filtering. However, for an effective use in image segmentation and shape decomposition, a clear interpretation of the spectral response regarding size and intensity scales is needed but lacking in current approaches. In this context, L^1 data fidelities are particularly helpful due to their interesting multi-scale properties such as contrast invariance. Hence, the novelty of this work is the combination of L^1-based multi-scale methods with nonlinear spectral decompositions. We compare L^1 with L^2 scale-space methods in view of spectral image representation and decomposition. We show that the contrast invariant multi-scale behavior of $L^1 - TV$ promotes sparsity in the spectral response providing more informative decompositions. We provide a numerical method and analyze synthetic and biomedical images at which decomposition leads to improved segmentation.

2.1 Introduction

In imaging science, the solution of inverse problems is often addressed by the modeling and analysis of variational methods of the form:

$$\min_u \frac{1}{p} \|u - f\|_{L^p}^p + \alpha J(u) \tag{2.1}$$

where f denotes a noisy signal, u is a desired image function defined on Ω in $\mathbb{R}^2$. The data fidelity measures the residual in the L^p norm and J is a regularization functional that has a weighting parameter $\alpha \geq 0$. In this paper, we concentrate on image denoising methods with convex and one-homogeneous regularization functionals J which can address image decomposition and segmentation adequately. More specifically, we focus on total variation regularization $J(u) = TV(u)$ and norms with $p = 1$ versus $p = 2$. One option to generate a scale-space method is a gradient flow, based on the functional (2.1) with initial condition $u(0, x) = f(x)$ and subdifferential inclusions forming the doubly nonlinear evolution equation:

$$0 \in \partial \|\partial_t u(t, x)\|_{L^p} + \partial J(u(t, x)). \tag{2.2}$$

In the case of $p = 2$, this simplifies to well-known gradient flows of the form:

$$\partial_t u = -q(u) \quad \text{with } q \in \partial J(u). \tag{2.3}$$

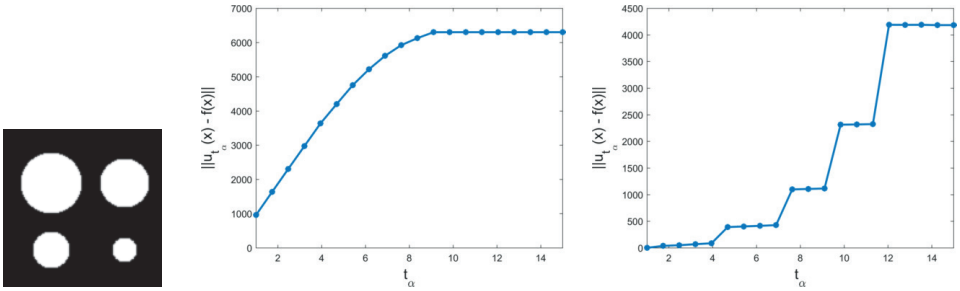
The regularization parameter α is now hidden in the time dependency. Corresponding inverse scale space gradient flows can be constructed via Bregman distances [3, 19]. The analysis of linear eigenvalue problems and spectral decompositions, e.g. via the Fourier transform, is a well-known and widely used theory in the fields of signal, image and graph-based data processing. Due to the continuing success of nonlinear regularization functionals in imaging, there is strong interest in generalizing spectral theory to the nonlinear case. The general idea is to examine solutions to the nonlinear eigenvalue problem:

$$\lambda u \in \partial J(u). \tag{2.4}$$

In [13, 14], Gilboa introduced the idea of nonlinear spectral decompositions for the TV transform. By transferring solutions (eigenfunctions) of (2.4) to sparse peaks in a spectral domain, the idea of advanced filters, suppressing or enhancing image components similar to the Fourier transform, came about. This concept was studied for one-homogenous functionals [4, 5, 15] and scale-space flows of the form (2.3) with $p = 2$. In this way, a decomposition of the input signal f into significant components is possible while an exact representation of f can still be guaranteed.

For $p = 1$, the gradient flow in (2.2) is interesting and more challenging than for $p = 2$. Another way to obtain a forward scale-space is to construct a sequence of variational problems of the form (2.1) with increasing regularization parameter t_α replacing the fixed α. Here, the scale parameter t_α corresponds to the time variable t used in (2.2). In this paper, we will focus on this type of model particularly in view of $p = 1$. From pioneering works on nonlinear L^1 filtering [1, 9, 18, 22], it is well known that such variational reconstruction methods share interesting multi-scale properties including contrast invariance.

For this reason, L^1 data fidelities have been successfully used for advanced image reconstruction techniques [11, 21, 23], vector field estimation [24] and image decompositions regarding texture [1, 16]. The special multi-scale behavior becomes clear in Figure 2.1. In the $L^1 - TV$ case shown in (c), the contrast invariance leads to plateaus in the scale-space graph indicating an abrupt disappearance of TV-eigenshapes in the input image shown in (a). Motivated by [9, 12, 22] and nonlinear spectral methods [15, 25], the main goal of this work is to study L^1 versus L^2 in view of the sparsity of nonlinear spectral decompositions. Does L^1 imply sparsity and hence a more *informative spectral response*? Can we expect more *reliable image decompositions* via backtransformation facilitating improved image segmentation? What happens if complex shapes, e.g. nonconvex or *compositions of eigenfunctions*, are involved?



(a) (b) (c)

Figure 2.1: *Scale-spaces.* (a) input image f, (b) $L^2 - TV$, (c) $L^1 - TV$.

2.2 Modeling

In the following section we first give a short overview of the eigenshapes of the total variation functional and calibrable sets before we introduce the spectral framework for $L^2 - TV$ denoising in more detail. Finally, we show why a combination of the spectral framework with $L^1 - TV$ denoising seems promising and how the spectral framework can be adapted in this case.

2.2.1 Geometry: Eigenfunctions and Calibrable Sets

In the introduction in Figure 2.1 we have already seen an input image composed of TV eigenfunctions. A general geometric description of TV eigenfunctions is given in [2, 6, 7] in terms of calibrable sets, more precisely by convex sets which are Cheeger in themself. An indicator function $\chi_C(x)$ of a convex and connected set C with finite perimeter $Per(C)$, for which C admits

$$\operatorname*{ess\,sup}_{x \in \partial C} \kappa_C(x) \leq \frac{Per(C)}{|C|} \qquad (C \text{ is Cheeger in itself})$$

where $|C|$ denotes the area and κ_C the curvature of $\partial C \in C^{1,1}$, is a solution of (2.4) with unit norm and therefore an eigenfunction of TV. With this geometric interpretation of eigenfunctions for TV as Cheeger sets, the role of perimeter and volume is significant for contrast invariant image decompositions. In the convex case, Duval and collaborators [12] proved that exact solutions of the $L^1 - TV$ problem are given by an morphological opening followed by a simple test over the perimeter-area ratio. This fact was first published but not proved in [10]. For more complex shapes, formed by compositions of eigenfunctions, a better understanding of the L^1 scale-space flows is therefore a very promising direction.

2.2.2 $L^2 - TV$ based Spectral Analysis

The nonlinear spectral analysis framework was first introduced by Gilboa in [13, 14] for the total variation regularization functional. A forward scale-space was constructed via the TV flow (2.3). Here, the first iterate is the original data f which is then smoothed in every time-step such that increasingly fine scales are removed. This concept was later generalized to more general one-homogenous regularization functionals $J(u)$ in [4, 5, 15]. Moreover, it was shown that the forward scale-space can also be constructed via a variational approach by iteratively solving the ROF model (2.1) for $p = 2$ and $J(u) = TV(u)$ using increasing regularization parameters t_α. The idea of the spectral filtering approach is to transform the signal f into a spectral domain, where both filtering of certain scales and decomposition of f into significant signals is possible. Thus, eigenfunctions will be mapped onto a peak in the spectral domain. The shape of the eigenfunction is determined by the chosen regularization functional, for $J(u) = TV(u)$ the most prominent eigenfunction is a disc with radius r surrounding the origin. In [20], Strong and Chan analyzed how the solution of the ROF model behaves for increasing t_α if f is an eigenfunction. Thus, let $f(x) = c \cdot 1_{B_r(0)}(x)$ be a disc of constant height c and radius r surrounding the origin and with a background of zero. Then the solution of the $L^2 - TV$ model is given as:

$$u(t_\alpha, x) = \begin{cases} \left(c - \frac{2}{r}\, t_\alpha\right) \cdot 1_{B_r(0)}(x) & \text{if } 0 \leq t_\alpha < \frac{cr}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

That means that even in the noise-free case, the reconstructed solution $u(t_\alpha, x)$ never reaches the true value f and for increasing regularization parameter t_α the disc height decreases. Solutions are corrupted by a systematic contrast loss that is dependent on the regularization strength but also on the radius r and the height c of input data f. To transform such eigenfunctions to a peak in the spectral domain, Gilboa defined the spectral transform function $\phi(t, x)$ and the spectral response function $S(t)$ as:

$$\phi(t, x) = u_{tt}(t, x) \cdot t \quad \text{and} \quad S(t) = \|\phi(t, x)\|_{L^1}\,. \tag{2.6}$$

The definition of ϕ allows, under certain conditions, that the original signal f can be reconstructed via:

$$f(x) = \int_0^\infty \phi(t, x) dt + \bar{f}$$

where $\bar{f}$ is the mean of f. Filtered versions of f can be constructed by applying:

$$f_H(x) = \int_0^\infty H(t)\phi(t,x)dt + H(\infty)\bar{f}$$

where $H(t)$ is the filter function.

However, a disadvantage of the L^2 based spectral framework is ambiguity with respect to size and intensity scales. The method is not able to clearly differentiate size and intensity scales since the timepoint t_d at which a disc disappears and a peak occurs is determined by both values together.

2.2.3 $L^1 - TV$ based Spectral Analysis

In the following section we want to combine the idea of nonlinear spectral TV analysis with the L^1 denoising model:

$$\min_u \ \|u - f\|_{L^1} + t_\alpha TV(u). \tag{2.7}$$

As mentioned earlier, this model shows very interesting multi-scale decomposition behavior and seems therefore very suitable to be combined with the spectral approach. In [9], Chan and Esedoglu showed that the behavior of solutions of the $L^1 - TV$ model if f is the TV eigenfunction $f(x) = c \cdot 1_{B_r(0)}(x)$ is significantly different from the L^2 case. Here, the solution is given by:

$$u(t_\alpha, x) = \begin{cases} f & \text{if } 0 \le t_\alpha < \frac{r}{2} \\ c' \cdot f \text{ with } c' \in [0,1] & \text{if } t_\alpha = \frac{r}{2} \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

An interesting observation is that the solution $u(t_\alpha, x)$ is **not** dependent on the height c of the disc but only on the radius r. That means that the $L^1 - TV$ denoising approach is contrast-invariant and therefore highly suitable for decomposing a signal f based on size scales. Note that the solution of (2.7) is not unique; for f defined as above and $t_\alpha = \frac{r}{2}$ there exist, for example, an infinite number of solutions.

Since the gradient flow (2.2) is more challenging for $p = 1$ rather than $p = 2$ (due to non-smoothness of the data-term), we construct a forward scale-space by taking a variational approach; in other words, we solve (2.7) for increasing regularization parameters $t_{\alpha_1} < t_{\alpha_2} < \ldots < t_{\alpha_N}$. To transform an eigenfunction of $TV(u)$ onto a peak in the spectral domain, a suitable spectral transformation function is now given by:

$$\phi(t,x) = -u_t(t,x) \ \text{ fulfilling } \ f(x) = \int_0^\infty \phi(t,x)dt + \hat{c}. \tag{2.9}$$

Here, the first time derivative of $u(t,x)$ is in a distributional sense and the constant $\hat{c}$ is the median of f. Since the disc's height does not decrease in every time step but remains the constant c until it immediately decreases to 0, already the first time-derivative leads to a delta peak in the spectral transform function defined via:

$$S^2(t) = \langle \phi(t,x), f(x) \rangle. \tag{2.10}$$

This definition was first introduced by Burger and collaborators in [5] and leads to an analogue to Parseval's identity:

$$\|f\|^2 = \langle f, f\rangle = \int_0^\infty \langle \phi(t,x), f(x)\rangle \mathrm{d}t = \int_0^\infty S^2(t)\mathrm{d}t. \tag{2.11}$$

The signal can be filtered based on size via:

$$f_H(x) = \int_0^\infty H(t)\phi(t,x)dt + H(\infty)\hat{c} \tag{2.12}$$

where $H(t)$ is again the filter function. A segmentation of objects in a certain size range (nearly independent of their intensity; only objects with intensity $\hat{c}$ cannot be found) is given by:

$$f_{\mathrm{seg},H}(x) = \left(\int_0^\infty H(t)\phi(t,x)dt > 0\right). \tag{2.13}$$

Applications where this intensity-independent segmentation approach is especially helpful are described in section 2.4.

2.3 Numerical Approach

The following section introduces the numerical realization of the spectral $L^1 - TV$ framework presented previously. The main component is the numerical solution of the denoising problem. To find solutions of this purely primal nonlinear minimization problem (2.7), we make use of the first order primal dual algorithm proposed by Chambolle and Pock [8]. As the name already suggest, the minimization scheme works with a primal dual version of (2.7) given by:

$$\langle \nabla u, g\rangle + \|u - f\|_{L^1} - t_\alpha \delta_P(g) \longrightarrow \min_u \max_g \tag{2.14}$$

where $P = \{g : \|g\|_\infty \leq 1\}$ and $\delta_P(g)$ equals 0 if $g \in P$, and equals ∞ if $g \notin P$. We define $K(u) = \nabla u$, $F(u) = \|u\|_{L^1}$ and $G(u) = \|u - f\|_{L^1}$.

The minimization algorithm proposed by Chambolle and Pock consists of three update steps: the first step is a dual update using the resolvent operator of F^* and the second is a primal update using the resolvent operator of G. These are followed by a simple weighting step between the previous two primal iterates. See [8] for more details. The resolvent operators for G and F^* are presented in [8, chpt. 6.2]. To construct a forward scale space, we solve the $L^1 - TV$ denoising model with increasing regularization parameter $t_{\alpha_1} < t_{\alpha_2} < \ldots < t_{\alpha_N}$ and compute the spectral transform function $\phi(t,x)$ via backward-differences and the response function $S(t)$ based on these solutions. Both the resulting primal-dual algorithm to minimize (2.7) and the computation of the spectral functions are embodied in Algorithm 3. Note that to receive a very high degree of convergence we needed a very large number of iterations resulting in a long computational time. With fewer iterations, the algorithm did not converge completely and eigenfunctions lost contrast altough, according to (2.8), this should not be the case.

Algorithm 3 First-order primal-dual algorithm to solve (2.7).

Parameters: data f, reg. param. $0 < t_{\alpha_1} < t_{\alpha_2} < \ldots < t_{\alpha_N}$, $\tau, \sigma > 0$,
$\theta \in [0,1]$, $maxIts \in \mathbb{N}$
Initialization: $n = 0,\ u^0 = 0,\ p_0 := 0, \bar{u}^0 = u^0$

Iteration:

for $(i = 1 : N)$ **do**

1. Set $\alpha = t_{\alpha_i}$.

   while $(n < maxIts)$ **do**

   a) $g^{n+1} = \text{Proj}_{\{\{g:\|g\|_\infty \leq 1\}\}}(g^n + \sigma \nabla \bar{u}^n)$.

   b) $\text{arg}_u = u^n + \tau \nabla \cdot g$.

   c) $u^{n+1}(x) = \begin{cases} \text{arg}_u(x) - \frac{\tau}{\alpha} & \text{if} \quad \text{arg}_u(x) - f(x) > \frac{\tau}{\alpha} \\ \text{arg}_u(x) + \frac{\tau}{\alpha} & \text{if} \quad \text{arg}_u(x) - f(x) < -\frac{\tau}{\alpha} \\ f(x) & \text{if} \quad |\text{arg}_u(x) - f(x)| < \frac{\tau}{\alpha} \end{cases}$.

   d) $\bar{u}^{n+1} = u^{n+1} + \theta(u^{n+1} - u^n)$.

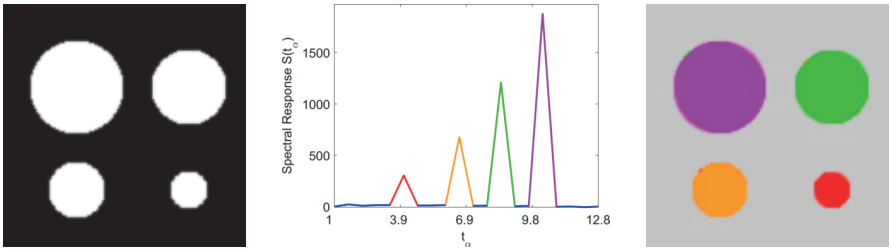
   e) Set $n = n + 1$.

   end while

2. Set $u(t_{\alpha_i}, x) = u^n$.
3. Set $\phi(t_{\alpha_i}, x) = u(t_{\alpha_{i-1}}, x) - u(t_{\alpha_i}, x)$.
4. Set $S(t_{\alpha_i}) = \langle \phi(t_{\alpha_i}, x), f(x) \rangle$.

end for

return $u(t_{\alpha_1}, x), \ldots, u(t_{\alpha_N}, x)$, ϕ, S.



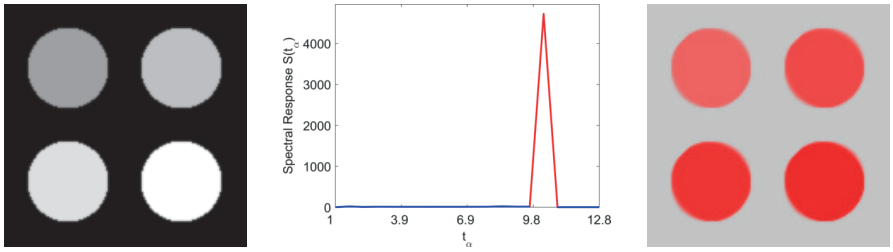
(a) Input Data. (b) Spectral Response. (c) Color-coded Reconstruction.

Figure 2.2: *Detection of size scales for eigenfunctions with constant intensity.* (a) shows the input data, (b) the spectral response function of the forward $L^1 - TV$ denoising approach. Every peak in S corresponds to a disc in the color-coded reconstruction in (c).

2.4 Results

In this section, we describe the main properties and advantages of the $L^1 - TV$ based spectral approach introduced above. We discuss both several synthetic experiments and real experiments from biological cell imaging and retina imaging. For some examples, the results are compared to results from the L^2 based spectral framework to illustrate the differences between both models. For all results, we applied the framework summarized in Algorithm 3 and cluster the spectral response function afterwards. In our experiments, we set $\tau = 0.2, \sigma = 0.625, \theta = 1$ and $maxIts = 50.000$. For synthetic datasets we used $N = 20$ linearly spaced t_{α_i} and increased N to 50 for the experimental datsets. After solving the variational model with increasing regularization parameter, we manually clustered S but comparable results can be achieved with common histogram thresholding methods such as Otsu's method, the Triangle method or methods designed for more than two classes. The different classes in the histogram are always visualized by different colors. Reconstruction of the filtered signal was performed via (2.12) where $H(t)$ was defined as an indicator function of the filtered time interval and the resulting signal was color-coded with the same color as in the spectral response and multiplied with the original gray values so that intensity changes in the input signal remain detectable.

Detection of Varying Eigenshapes: Size versus Intensity. In this set of experiments, we focus on circular objects that are all eigenfunction of the total variation functional. The aim of these experiments is to investigate which scales can be detected and reconstructed using our method and where the differences to an L^2 based approach are. Figure 2.2 shows an example with four discs in front of a uniform background, all showing the same intensity level. The corresponding spectral response function (b) clearly shows the four peaks each corresponding to one disc, as can be seen in the color-coded reconstruction in (c). Since L^1 is a purely size-based approach, it decomposes these eigenfunctions clearly. The opposite case is shown in Figure 2.3. Here, the input data is again composed of four discs but now



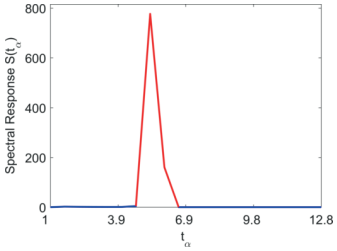
(a) Input Data. (b) Spectral Response. (c) Color-coded Reconstruction.

Figure 2.3: *Invariance w.r.t. to intensity scales.* (a) shows the input data consisting of four TV eigenfunctions of similar size but various intensities. The corresponding spectral response is shown in (b). All discs vanish at the same moment, independent of their contrast. (c) The reconstruction of the peak is shown in red.

with only one size and various intensity levels. Although an L^2 based approach would again show four peaks in the spectral response function due to its size/intensity ambiguity mentioned previously, we now see only one peak in the response (b). Filtering only the signal belonging to this red peak returns all four discs; see (c). This example clearly shows the contrast invariance of the L^1 data fidelity term which is a major difference between $p = 2$ and $p = 1$.

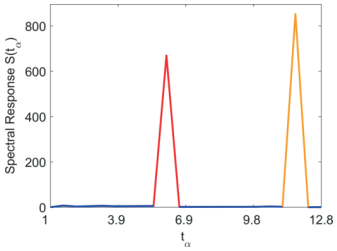
A direct comparison of both spectral frameworks is shown in Figures 2.4 and 2.5. Figure 2.4 shows as input data (a) two discs with different sizes and intensities. From a visual perspective, this would be clearly identified as two different scales but the $L^2 - TV$ denoising approach is not able to distinguish between the objects. The scale that this approach uses is always a mixture of the size scale (small for the disc on top and large for the other one) with the intensity scale (large for the one on top and small for the second disc) and therefore it can occur that they both end up with the same “medium" scale. In the spectral domain, they are represented by one peak (b) and can therefore not be reconstructed separately. However, this is different for the contrast invariant $L^1 - TV$ approach since it is purely size based. The spectral response function shows two easily separable peaks (d) that can be reconstructed one by one (e). A similar behavior can be seen for two discs on top of each other (see Figure 2.5 (a)). Although these two discs are represented by two separate peaks in both approaches, the L^2 based approach mixes them if they are on top of each other. In this case, the peaks become less sparse and apart from each other (b) and in the reconstruction of the small scales (c) we see some artifacts of the larger disc (bluish ring around red disc). This is not the case for L^1. Both peaks are clearly separable (e) and the reconstruction gives a clear separation of both discs. The larger red circle in (f) is just a discretization artifact.

(a) Input Data.



(b) Spectral Response of $L^2 - TV$ denoising.

(c) Color-coded Reconstruction (L^2).



(d) Spectral Response of $L^1 - TV$ denoising.

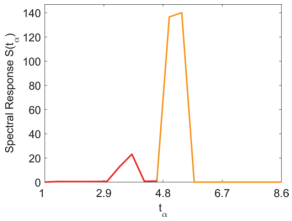
(e) Color-coded Reconstruction (L^1).

Figure 2.4: *Detection of size and intensity scale mixtures.* (a) shows the input data with two eigenfunctions with differing sizes and intensities. Results of an L^2 based spectral approach are shown in (b) and (c). The method is not able to separate the objects. (d) and (e) show the results of the contrast invariant L^1 based spectral analysis. The two discs are clearly separable based on size.

Segmentation of Experimental Cell Data. In Figure 2.6, we present a dataset that was also used in [25]. The experimental dataset (a) shows a fluorescence microscopy image of Circulating Tumor Cells. Here, the goal is to reliably segment all cells although they differ much in size and intensity. A multi-scale segmentation approach was presented in [25], see (d), but due to the intensity dependency of this approach, the method was not able to detect those cells that are very dim (highlighted with red boxes). Since our new L^1 based spectral approach ignores intensity differences, the method also finds the very dim cells. When reconstructing the orange part of the spectral response function (b) and using the thresholding formula (2.13), we obtain a segmentation that contains all four cells highlighted with a red box.

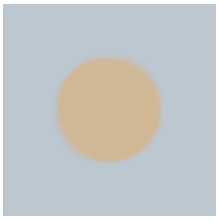
Experimental Data of Network Structures. In Figure 2.7 (a), a manually segmented blood vessel network taken from the STARE dataset [17] is shown. A problem often occurring in retinal blood vessel segmentation is that small vessels are also very dim and therefore even more challenging to detect. In (b), we added an intensity bias to the original data to test whether this influences our segmentation/clustering approach or not. In (c)-(f), the

(a) Input Data.

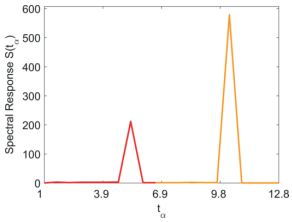


(b) Spectral Response of $L^2 - TV$ denoising.

(c) Color-coded Reconstruction of red peak (L^2).

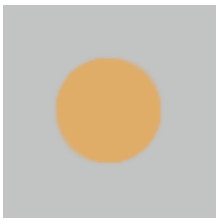


(d) Color-coded Reconstruction of orange peak (L^2).



(e) Spectral Response of $L^1 - TV$ denoising.

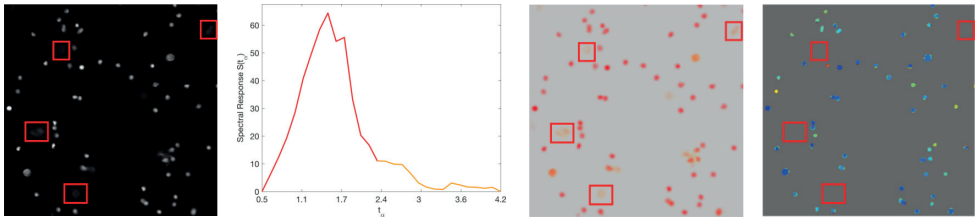
(f) Color-coded Reconstruction of red peak (L^1).



(g) Color-coded Reconstruction of orange peak (L^1).

Figure 2.5: *Overlapping mixtures of size and intensity scales.* (a) shows the input data with two discs with different size and intensity on top of each other. (b) - (d) shows the spectral response and the reconstructions of both peaks using an L^2 dataterm and (e) - (g) for an L^1 dataterm. In (c) we see artifacts of the larger disc already appearing at the fine scales while (f) and (g) clearly separate the two discs.

spectral response functions and reconstructions for L^2 denoising are shown. We see that there are fewer clear peaks in the spectral response function, especially in the case of intensity biased input. Another problem that we observe is that the vessels are not removed while retaining their original shape but are reshaped to more circular objects. This leads to a mixing of all scales in the reconstruction and the dim appearance of the visualization. For L^1 based denoising, the spectral responses are much sparser in (h) and (j) and therefore easier to cluster. The reconstructions (g) and (j) are also much easier to interpret since the vessels are not reshaped but removed in one step based on their diameter. Thus, for both networks (even for the one with an intensity bias) we see a clear clustering of blood vessels based on size. The difference between both methods is clearly shown in the magnified box. While in the top row the largest scale is just a turquoise shadow around the thickest blood vessels, the blood vessel is clearly reconstructed in the bottom row.



(a) Input Data. (b) Spectral Response. (c) Color-coded Reconstruction. (d) Segmentation using the Bregman-CV.

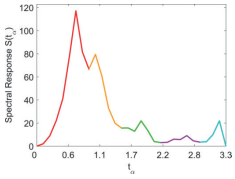
Figure 2.6: *Segmentation of cells of multiple sizes and intensities.* (a) shows an experimental dataset of tumor cells under a fluorescent microscope; (b) the resulting spectral response function; and (c) the color-coded reconstruction. A segmentation result taken from [25] is given in (d). Wee see that both methods can be used to obtain a multi-scale segmentation but very dim objects (red boxes) might be lost when taking an intensity based approach as in (d). In taking our purely size-based approach the dim cells are also found.

2.5 Conclusion and Outlook

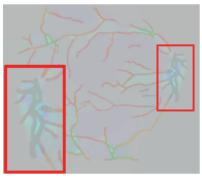
In this paper, we have described the study of contrast-invariant L^1 data fidelities for variational multi-scale methods in combination with nonlinear spectral image analysis. We have shown that the contrast invariance results in an improved sparsity of spectral responses. In comparison to standard L^2, this allows a more informative spectral image representation to be obtained. We presented a model, an efficient algorithm and numerical results. In the particular case of experimental data sets that have complex shapes and strong intensity variations of objects or the background, our method outperforms the current standard method for nonlinear spectral decomposition. For future studies, it will be important to extend the ideas to nonlocal graph-based problems such as L^1 with nonlocal TV, and further analyze the relationship of the doubly nonlinear scale-space flow to the proposed scale-space procedure.

(a) Unbiased Input.

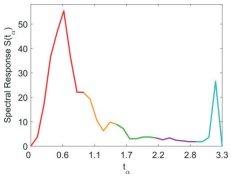
(b) Biased Input.



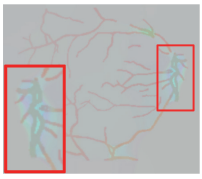
(c) Spectral Response (L^2, Unbiased).



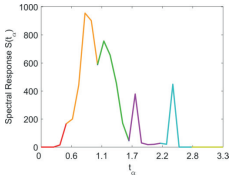
(d) Color-coded Reconstruction (L^2, Unbiased).



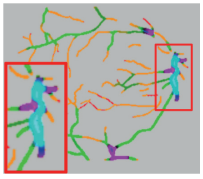
(e) Spectral Response (L^2, Biased).



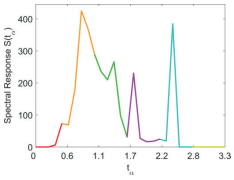
(f) Color-coded Reconstruction (L^2, Biased)



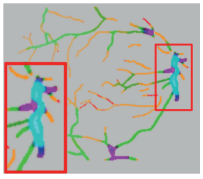
(g) Spectral Response (L^1, Unbiased).



(h) Color-coded Reconstruction (L^1, Unbiased).



(i) Spectral Response (L^1, Biased).



(j) Color-coded Reconstruction (L^1, Biased)

Figure 2.7: *Decomposition of network structures without and with intensity bias.* (a) shows the binary input network of blood vessel [17] and (b) the same network with an intensity bias. (c)-(f) show spectral response and reconstruction results using L^2 and the unbiased (c)-(d) resp. biased data (e)-(f). In (g)-(j), the results with L^1 are presented. Reconstructions (d) and (f) appear very dim since vessels are not removed in one step but are reshaped over time, resulting in a mix of (colored) scales. This is more clear in the magnification. The turquoise part in (d) and (f) is no longer vessel-shaped but more a roundish shadow around the original shape while the structure remains unchanged for L^1 based reconstruction (h) and (j). This network structure is also not affected by an intensity bias.

Bibliography

[1] J.-F. Aujol, G. Gilboa, T. F. Chan, and S. J. Osher. Structure-Texture Image Decomposition – Modeling, Algorithms, and Parameter Selection. *International Journal of Computer Vision*, 67(1):111–136, 2006.

[2] G. Bellettini, V. Caselles, and M. Novaga. The Total Variation Flow in RN. *Journal of Differential Equations*, 184(2):475–525, 2002.

[3] C. Brune, A. Sawatzky, and M. Burger. Primal and Dual Bregman Methods with Application to Optical Nanoscopy. *International Journal of Computer Vision*, 92(2):211–229, 2011.

[4] M. Burger, L. Eckardt, G. Gilboa, and M. Moeller. Spectral Representations of One-Homogeneous Functionals. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 16–27. Springer, 2015.

[5] M. Burger, G. Gilboa, M. Moeller, L. Eckardt, and D. Cremers. Spectral Decompositions Using One-Homogeneous Functionals. *SIAM Journal on Imaging Sciences*, 9(3):1374–1408, 2016.

[6] V. Caselles, A. Chambolle, and M. Novaga. Uniqueness of the Cheeger Set of a Convex Body. *Pacific Journal of Mathematics*, 232(1):77–90, 2007.

[7] A. Chambolle, V. Duval, G. Peyré, and C. Poon. Geometric Properties of Solutions to the Total Variation Denoising Problem. *Inverse Problems*, 33(1):015002, 2016.

[8] A. Chambolle and T. Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[9] T. F. Chan and S. Esedoglu. Aspects of Total Variation Regularized L1 Function Approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837, 2005.

[10] J. Darbon. Total Variation Minimization with L1 Data Fidelity as a Contrast Invariant Filter. In *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pages 221–226, 2005.

[11] Y. Dong, M. Hintermüller, and M. Neri. An Efficient Primal-Dual Method for L1 TV Image Restoration. *SIAM Journal on Imaging Sciences*, 2(4):1168–1189, 2009.

[12] V. Duval, J.-F. Aujol, and Y. Gousseau. The TVL1 Model: A Geometric Point of View. *Multiscale Modeling & Simulation*, 8(1):154–189, 2009.

[13] G. Gilboa. A Spectral Approach to Total Variation. In A. Kuijper, K. Bredies, T. Pock, and H. Bischof, editors, *Scale Space and Variational Methods in Computer Vision*, pages 36–47. Springer, 2013.

[14] G. Gilboa. A Total Variation Spectral Framework for Scale and Texture Analysis. *SIAM Journal on Imaging Sciences*, 7(4):1937–1961, jan 2014.

[15] G. Gilboa, M. Moeller, and M. Burger. Nonlinear Spectral Analysis via One-Homogeneous Functionals: Overview and Future Prospects. *Journal of Mathematical Imaging and Vision*, 56(2):300–319, 2016.

[16] A. Haddad. Texture Separation BV-G and BV-L1 Models. *Multiscale Modeling & Simulation*, 6(1):273–286, jan 2007.

[17] A. Hoover, V. Kouznetsova, and M. Goldbaum. Locating Blood Vessels in Retinal Images by Piecewise Threshold Probing of a Matched Filter Response. *IEEE Transactions on Medical imaging*, 19(3):203–210, 2000.

[18] M. Nikolova. Minimizers of Cost-Functions Involving Nonsmooth Data-Fidelity Terms. Application to the Processing of Futliers. *SIAM Journal on Imaging Sciences*, 40(3):965–994, 2002.

[19] S. J. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An Iterative Regularization Method for Total Variation-Based Image Restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[20] D. Strong and T. F. Chan. Edge-Preserving and Scale-Dependent Properties of Total Variation Regularization. *Inverse Problems*, 19(6):S165, 2003.

[21] C. Wu, J. Zhang, and X.-C. Tai. Augmented Lagrangian Method for Total Variation Restoration with Non-Quadratic Fidelity. *Inverse Problems and Imaging*, 5(1):237–261, 2011.

[22] W. Yin, D. Goldfarb, and S. J. Osher. The Total Variation Regularized L1 Model for Multiscale Decomposition. *Multiscale Modeling & Simulation*, 6(1):190–211, 2007.

[23] J. Yuan, J. Shi, and X.-C. Tai. A Convex and Exact Approach to Discrete Constrained TV-L1 Image Approximation. *East Asian Journal on Applied Mathematics*, 1(2):172–186, 2011.

[24] C. Zach, T. Pock, and H. Bischof. A Duality Based Approach for Realtime TV-L1 Optical Flow. In *Joint Pattern Recognition Symposium*, pages 214–223. Springer, 2007.

[25] L. L. Zeune, G. van Dalum, L. W. Terstappen, S. A. van Gils, and C. Brune. Multiscale Segmentation via Bregman Distances and Nonlinear Spectral Analysis. *SIAM Journal on Imaging Sciences*, 10(1):111–146, 2017.

Chapter 3

Quantifying HER-2 Expression on Circulating Tumor Cells by ACCEPT

Leonie L. Zeune, Guus van Dalum, Charles Decraene, Charlotte Proudhon, Tanja Fehm, Hans Neubauer, Brigitte Rack, Marianna Alunni-Fabbroni, Leon W.M.M. Terstappen, Stephan A. van Gils and Christoph Brune

PloS one 12.10 (2017): e0186562

Abstract

Circulating tumor cells (CTCs) isolated from blood can be probed for the expression of treatment targets. Immunofluorescence is often used for both the enumeration of CTCs and the determination of protein expression levels related to treatment targets. Accurate and reproducible assessment of such treatment target expression levels is essential for their use in the clinic. To enable this, an open source image analysis program named ACCEPT was developed in the EU-FP7 CTCTrap and CANCER-ID programs. Here its application is shown on a retrospective cohort of 132 metastatic breast cancer patients from which blood samples were processed by CellSearch® and stained for HER-2 expression as additional marker. Images were digitally stored and reviewers identified a total of 4084 CTCs. CTC's HER-2 expression was determined in the thumbnail images by ACCEPT. 150 of these images were selected and sent to six independent investigators to score the HER-2 expression with and without ACCEPT. Concordance rate of the operators' scoring results for HER-2 on CTCs was 30% and could be increased using the ACCEPT tool to 51%. Automated assessment of HER-2 expression by ACCEPT on 4084 CTCs of 132 patients showed 8 (6.1%) patients with all CTCs express-

ing HER-2, 14 (10.6%) patients with no CTCs expressing HER-2 and 110 (83.3%) patients with CTCs showing a varying HER-2 expression level. In total 1576 CTCs were determined HER-2 positive. We conclude that the use of image analysis enables a more reproducible quantification of treatment targets on CTCs and leads the way to fully automated and reproducible approaches.

3.1 Introduction

Peripheral blood tumor load represented by CTCs is associated with poor outcome in cancer patients [10, 12, 13, 19, 24]. The availability of CTCs allows for the assessment of treatment targets and opens the avenue to provide CTC-based therapy to the patient. The ability to detect treatment targets on CTCs has been demonstrated in a variety of studies [4, 8, 9, 16, 17, 18, 23]. Before this information can be used in the clinic it is imperative that such a target can be reproducibly and consistently quantified on the CTCs at different clinical sites. Although the majority of multicenter studies have been performed with the FDA cleared CellSearch® system, in recent years many systems have been introduced to detect and isolate CTCs [1, 5, 20]. The lack of a unified approach to designate a cell as a CTC, and to determine whether or not a CTC expresses a treatment target, leads to large differences in reported CTC numbers and positivity rates for potential therapeutic targets such as HER-2 between various studies urging the need for standardization. To address this issue a CTC image analysis algorithm for identification and characterization of CTCs is being developed in the EU funded CANCER-ID & CTCTrap programs. Here we introduce the first version of the Open Source program named ACCEPT (Automated CTC Classification Enumeration and PhenoTyping) that allows for the quantification of treatment targets on annotated CTCs. ACCEPT is a toolbox in which a novel efficient parameter-free multi-scale segmentation method is used to identify objects in images captured by several CTC platforms [25]. Here we investigate a retrospective set of images generated from CTCs detected in metastatic breast cancer patients' blood by the CellSearch® system in which the expression of HER-2 is assessed. Images generated by the CellSearch® system were used as this system is standardized and in use in multiple centers. But any annotated set of tiff images can be analyzed using the toolbox, allowing for the quantification of markers independent of the CTC enrichment platform that is used. The toolbox is available for use at https://github.com/LeonieZ/ACCEPT.

3.2 Materials and Methods

3.2.1 CTC Enumeration and HER-2 Assessment

The CellSearch® system (Menarini Silicon Biosystems Inc, Huntingdon Valley, PA, USA) was used to enumerate CTCs and to assess their relative HER-2 expression. The cells enriched from 7.5 mL of blood by EpCAM-expression were labeled with phycoerythrin (PE) conjugated antibodies directed against epithelial cell specific cytokeratins (CKs), with an allophycocyanin conjugated (APC) antibody directed against leukocyte specific CD45, and with a Flu-

orescein (FITC) conjugated antibody (HER81) directed against HER-2. Additionally, cells were stained with the nuclear dye 4,6-diamidino-2-phenylindole (DAPI) to identify the nucleus. The CellTracks Analyzer II® (Menarini Silicon Biosystems Inc) was used to acquire digital images of the four different fluorescent dyes using a 12-bit camera that are transformed and store as 8-bit images during archiving. Trained operators reviewed thumbnail images generated by the CellTracks Analyzer II® to count and determine the HER-2 expression of the CTCs according to the manufacturer's instructions.

3.2.2 Cell Lines

Breast carcinoma cell lines SKBR-3, MDA-MB 453 and MDA-MB 231 were obtained from ATCC (Manassa, VA, USA) and cultured in DMEM (Gibco, Life Technologies, Waltham, MA, USA) containing 2 mM L-glutamine (G7513, Sigma-Aldrich), 100 U/mL penicillin and 100 μg/mL streptomycin (P4333, Sigma-Aldrich) and 10% FBS (F4135, Sigma-Aldrich) at 37°C in a humidified 5% CO2 incubator. Cells were trypsinized at about 80% confluence with 0.05% Trypsin-EDTA (1X) with Phenol Red (Gibco, Life Technologies). HER-2 quantification was performed with flow cytometry using the QuantiBRITE® PE quantification kit (BD biosciences, San Jose, California) using a previously published protocol [21]. Peripheral blood of healthy donors was spiked with either 1500 SKBR-3, 500 MDA-MB 453 or 500 MDA-MB 231 cells and processed with the CellSearch® system.

3.2.3 Patients

In this study CTC images generated by the CellSearch® system (Menarini Silicon Biosystems Inc) from 132 patients with metastatic breast cancer were used. 80 patients (36 recruited in the Department of Gynecology and Obstetrics - Ludwig-Maximilians-University, Munich and 44 in the Department of Gynecology and Obstetrics - Heinrich-Heine-University, Duesseldorf) were enrolled in the Detect III study (NCT01619111). The other 52 patients were enrolled in the BEVERLY02 study from Paris (NCT00717405). All patient identifying information is maintained at the clinical sites and no access to this information was available for our analysis. Approval for HER-2 CTC analysis was obtained by the independent ethics committee for both, the Detect III study (Ethical Committee of the Heinrich-Heine University Duesseldorf) and the BEVERLY02 study (Comité de Protection des Personnes Sud Méditerrané I), and all patients provided written informed consent.

3.2.4 Image Analysis

The images were reanalyzed using the newly developed image analysis toolbox for CTC analysis ACCEPT (Automated CTC Classification Enumeration and PhenoTyping). The software is written in MATLAB® 2016a (Mathworks, Natick, MA). Next to the source code a compiled standalone executable version, that includes the MATLAB® Runtime and allows royalty-free deployment for users who do not need MATLAB®, is also available. In ACCEPT objects of interest are segmented based on shape contours and intensity. The underlying algorithm

to automatically detect multiple objects with different scales is described in detail elsewhere [6, 7, 25]. The main tool of the ACCEPT toolbox is a fully automated detection and classification approach for blood samples but in this paper we introduce a second tool, called the Marker Characterization tool, that reproducibly evaluates the marker expression of prescored cells and can aid users to manually score or quantify cells and their fluorescent signals. The thumbnail images of all 4084 manually scored CTCs out of 132 patients were loaded into the ACCEPT algorithm. The images that are stored as 8-bit images by the CellSearch® system with an image brightness automatically adjusted relative to the brightest pixel in the channel, are rescaled before processing to their true intensity values by extracting the original minimum and maximum value in each channel from the accompanying tiff header. Afterwards the outline of all events present in these images was automatically detected in each channel. Based on these contours, we extracted seven different measurements per object and per fluorescent channel, i.e. eccentricity (circularity measure), perimeter, mean intensity, maximum intensity, size, standard deviation of the intensity, mass (sum of all intensity values) and perimeter2area (circularity measure). Moreover, we extracted the relative overlay of the signals in the DAPI and CD45 channels and in the DAPI and CK channels. The thumbnail images together with the extracted contour and measurements are presented to the operator to determine the HER-2 expression.

3.3 Results

3.3.1 Visualization of CTCs by ACCEPT

In Figure 3.1 the visualization of CTCs in the ACCEPT toolbox is shown. Some sample information is presented on the top left with a scaled overview image of the sample next to it. The main components of the visualization window are the thumbnail gallery on the left and three scatter plots on the right. The thumbnail gallery presents an overlay image of the first three fluorescent markers (CD45, DAPI, CK) and next to it a thumbnail image of every single fluorescent channel. While the overlay image is scaled, the single channels present the full, unscaled intensity range to prevent misinterpretation of the signal. For the three marked cells a scaled version is depicted below the visualization window. While the scaled visualization gives the impression that all three cells have a similar expression, a precise segmentation of the signal as done in ACCEPT (indicated by the red contour) shows that their mean intensity does differ. The standard deviation of the background signal is in this case 6.6 on average. A two-sided t-test shows that the intensity difference between thumbnail one and three as well as between two and three is statistically meaningful. In the unscaled image, interpretation faults like this are prevented. Very dim signals (like the HER-2 expression of the three selected cells) are difficult to see, but the red contour shows that a signal is presented and the dots in the scatter plots on the right visualize the extracted measurements such as the mean intensity. In the scatter plots of Figure 3.1 we plotted the CK versus DAPI mean intensity, CK versus CD45 mean intensity and CK versus HER-2 mean intensity. Moreover, right clicking on an image in the thumbnail gallery opens a scaled visualization.

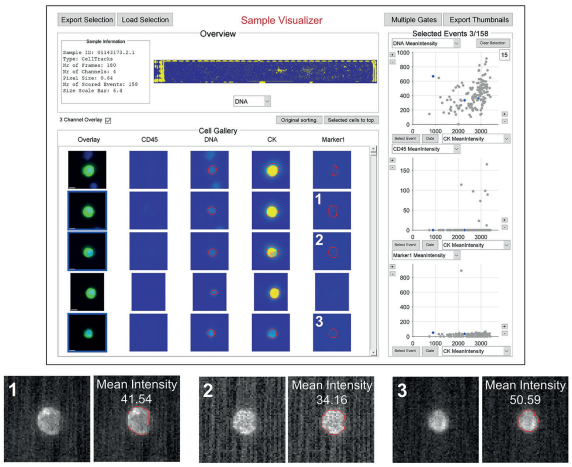
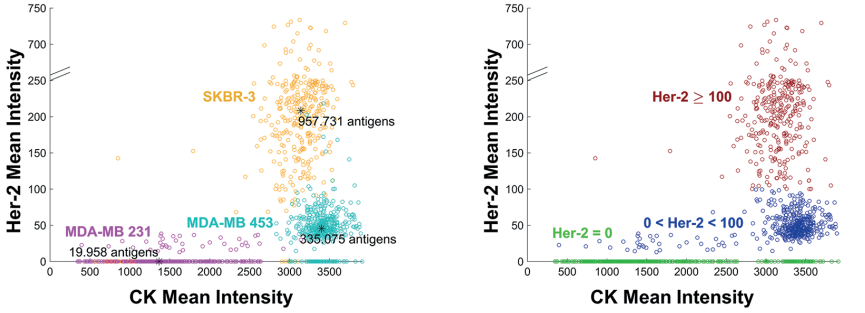


Figure 3.1: *Sample Visualizer of ACCEPT.* In the scatter plots 3 of the 158 objects are depicted blue and the corresponding thumbnail images are highlighted. In this example "Marker1" represents signals for HER-2. The corresponding HER-2 images are shown below the Sample Visualizer, in the right image the red line indicates the boundary detected by ACCEPT of the identified CTC and the number indicates the median value of the HER-2 staining within this boundary. Size bar in overlay: 6.4 μm

3

3.3.2 HER-2 Expression on Breast Cancer Cell Lines

To define a threshold for positive HER-2 expression, we evaluated the HER-2 mean intensity of the three breast cancer cell lines SKBR-3 (3+), MDA-MB 453 (2+) and MDA-MB 231 (0 or 1+). Figure 3.2 (a) depicts the scatter plot of the CK mean intensities versus the HER-2 mean intensities for each cell line. Based on the HER-2 expression shown in panel A we defined 2 thresholds to distinguish between HER-2 negativity (0 or 1+), dim HER-2 expression (2+) and bright HER-2 (3+) expression: a dim HER-2 expression has a mean intensity between 0 and 100 and a bright HER-2 expression has a mean intensity above 100. With these thresholds 87.15% of the cells were correctly classified (Figure 3.2 (b)). Note that a cutoff of 0 is sufficient to distinguish between HER-2 negative and dim HER-2 expression since a separating gap between these two classes is automatically constructed by our segmentation method. Yet we see that overall the HER-2 expression is very low; the maximal possible intensity is 4095 while the highest measured mean intensity in the 3+ positive cell line is around 1000. To ensure that the measured signal intensity correlates to the number of antigens we compared the median intensity value and coefficient of variation (CV) for each of the three cell lines to the number of HER-2 antigens determined by measuring their mean expression levels by flow cytometer using the BD Quantibrite™ Beads PE Fluorescence Quantitation Kit (BD Biosciences, San Jose, CA, USA). The average expression of HER-2 on the SKBR-3 cells was 957.731 (CV 78.4%) HER-2 antigens, MDA-MB 453 had 335.075 (CV 62.9%) HER-2 antigens and MDA-MB 231 expressed 19.958 (CV 108.1%) HER-2 antigens. This is in line with literature [22]. These values result in a linear correlation (see Supplementary Figure 3.1) and show that



(a) (b)

Figure 3.2: *Cytokeratin and HER-2 mean intensities of cells in breast cancer cell lines MDA-MB 231, MDA-MB 453 and SKBR-3.* Panel (a), 373 MDA-MB 231 cells (magenta), 496 MDA-MB 453 cells (cyan), 361 SKBR-3 cells (orange). Average number of HER-2 antigens included for each cell line. Panel (b) classification of the MDA-MB 231, MDA-MB 453 and SKBR-3 into 428 negative HER-2 (green), 462 dim HER-2 (blue) and 340 bright HER-2 (red) expressing cells identified by cluster analysis.

the measured mean intensity is a valid measure for the HER-2 expression. This was also previously shown in [14]. The authors showed that the HER-2 signal of the cells found using this assay relates to the gene copy number in their cohort.

3.3.3 HER-2 Expression of CTCs in Metastatic Breast Cancer Patients

We evaluated the HER-2 expression of 4084 CTCs from 132 metastatic breast cancer patients and applied the same thresholds as with the cell lines in Figure 3.2. Figure 3.3 shows a scatter plot with the CK mean intensity (x-axis) versus the HER-2 mean intensity (y-axis). Again, all CTCs not expressing HER-2 are depicted in green (61.4%), CTCs with a dim HER-2 expression between 1 and 99 are labeled blue (36.2%) and those with a bright HER-2 mean expression larger or equal to 100 are colored in red (2.4% of all CTCs).

3.3.4 Heterogeneity of HER-2 Expression per Patient

For each of the 132 patients we evaluated the heterogeneity of HER-2 expression in CTCs. The results are shown in Figure 3.4 and are sorted from left to right according to decreasing percentages of CTCs that are HER-2 dim or bright (HER-2 $>$ 0). Again, HER-2 negative CTCs are colored in green, dim HER-2 in blue and bright HER-2 in red. In 8 (6.1%) patients all CTCs expressed HER-2, in 14 (10.6%) patients none of the CTCs expressed HER-2 and in 110 (83.3%) patients various HER-2 expression levels on CTCs were observed. Thus, for most patients not all CTCs express or lack HER-2, most patients rather present CTCs with varying levels of HER-2 signal.

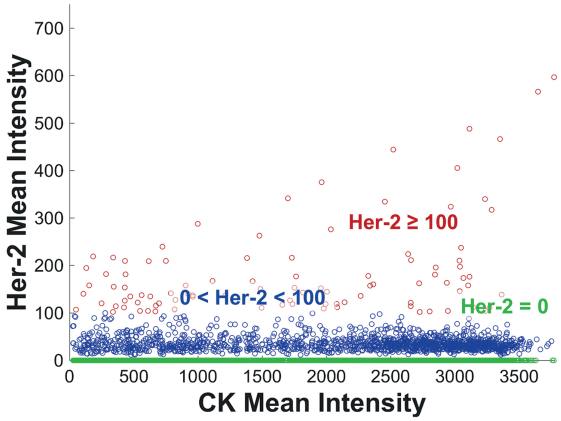


Figure 3.3: *Expression of Cytokeratin and HER-2 on 4084 CTCs in 132 breast cancer patients.* HER-2 negative CTCs are colored in green, dim HER-2 (0 < HER-2 < 100) in blue and bright HER-2 (>100) in red.

3.3.5 HER-2 expression of CTCs Assessed by Operators and ACCEPT

To assess whether the defined thresholds correlate with manual scoring, we compared for each of the 132 samples the number of automatically scored HER-2 positive CTCs (HER-2 mean intensity > 0) with the number of CTCs that were manually scored as HER-2 positive (2+ or 3+) by the three clinical sites. The results are shown in Figure 3.5. We observe a good agreement between the scores of sites 2 and 3, yet on average they manually scored more cells as HER-2 positive than our automated procedure. Nevertheless, there is a reasonably high Pearson correlation between the manual and automatic scores for site 2 (coefficient r=0.82) and site 3 (coefficient r=0.98). Operators of site 1 were much stricter with their definition of HER-2 positivity and scored fewer cells as positive compared to the automated approach. This results in a much lower correlation coefficient (r=0.51). To compare the number of manually and automatically scored cells to a “ground truth” solution, further experiments are necessary to obtain images of cells where the HER-2 expression is known. Yet, in this work, we concentrate on retrospectively studying samples that were investigated before. The main goal of the Marker Characterization tool of the ACCEPT toolbox is to reproducibly quantify manually scored cells and aid users to unify their scoring results. Yet, any approach that is still to some part manual will not give fully reproducible and unified results. This highly motivates the use of a fully automated approach, as it will be possible in another tool of the ACCEPT software, especially in multi-center studies to unify the scoring results.

3.3.6 Influence of Quantitative Display of HER-2 Expression

In Figure 3.1 we have shown the difference between the traditionally scored visualization of fluorescent intensity (three examples at the bottom) and the visualization we used in the ACCEPT tool. Since it is difficult to evaluate the intensity of a fluorescent signal based on

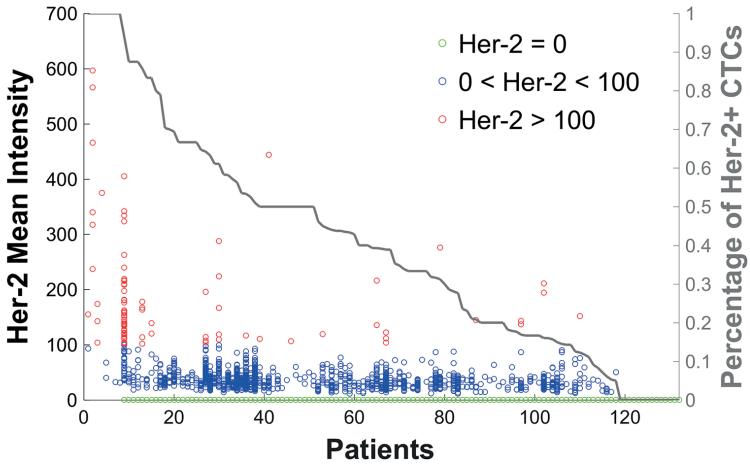


Figure 3.4: *HER-2 expression in CTCs in each of the 132 patients.* Patients are sorted according to the percentage of HER-2+ CTCs using a threshold of (>0). HER-2 negative CTCs are colored in green, dim HER-2 in blue (0 < HER-2 < 100) and bright HER-2 (>100) CTCs red.

a scaled image, this way of visualizing data could be a major cause for the high inter-user variability observed in manual scores. Therefore, we developed a tool to allow the operators to rate the HER-2 expression with and without the quantitative display of HER-2 expression in the form of a scatterplot. To test if the concordance between different operators can be increased when our tool is used for scoring, a total of 150 randomly chosen images of CTCs were sent to six different investigators for scoring HER-2 positivity. The results are shown in Figure 3.6. For scaled visualization images in 30 percent of the cases all investigators agreed with a major part of agreement on positivity (red box) although a large part of the cells had nearly no HER-2 expression by quantitative assessment of HER-2 mean intensities. With the quantitative display in ACCEPT this percentage of full agreement can be increased to over 50 percent (right graph). Here the investigators agreed mostly on HER-2 negativity (blue box). Thus, the scaled visualization has a high sensitivity but also a lot disagreement while the quantitative display leads to more true negatives and therefore a higher specificity and less disagreement. To investigate which cells have the highest probability to result in disagreement of reviewers, we compared the Cytokeratin and HER-2 mean intensity of each of the 150 cells to the number of reviewers that agree on their HER-2 status. The results for both, the scaled and the ACCEPT visualization, are shown in Supplementary Figure 3.2. We see that the quantitative display significantly decreases the spread of user disagreement. Yet, the Marker Characterization tool can still only help users to make their decision therefore there are still 49% where at least one reviewer has a different opinion than the other ones. This illustrates the need for either automated classification of CTCs as HER-2 positive CTCs or classification of HER-2 positive CTCs by using the ACCEPT quantitative displays.

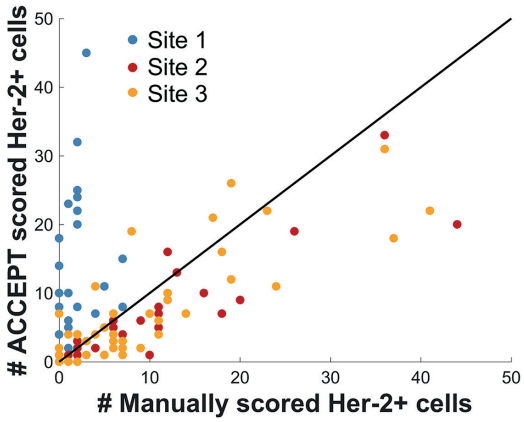


Figure 3.5: *Number of manually scored HER-2+ cells versus number of automatically scored HER-2+ cells using a threshold of* > 0. Samples of 132 patients were investigated.

3.4 Discussion

Since treatment options for cancer patients are increasing rapidly and concurrently there is a growing demand for biomarkers that can predict the most effective therapy. Availability of tumor material before the start of a new line of therapy is imperative to determine the presence of treatment targets to eradicate the tumor. However, a tumor or its metastasis might not always be accessible for biopsy. Therefore, isolation of tumor cells from blood represents a unique opportunity to obtain tumor material providing its latest genotypic and phenotypic fingerprint. After demonstration that tumor cells can be reproducibly isolated from blood and that their presence is related to poor clinical outcome [2] many new technologies have been introduced to detect and isolate CTCs [1, 3, 5, 20]. The lack of an automated unified approach to designate a cell as a CTC and determine what is and what is not expressed is however impeding progress of the field. To address this issue, an open source image analysis program ACCEPT is being developed in a consorted European effort enabling comparisons between platforms and providing accurate and reproducible information. Here we introduce the first applications of ACCEPT and demonstrate the ability to extract relative expression of antigens expressed by CTCs. In addition, we show that the use of scatterplots and expression levels of the antigens next to the fluorescence images helps to reduce inter reader variability in terms of scoring of HER-2 positivity. Increased concordance in the expression of targetable biomarkers will optimize the performance and results of clinical multicenter studies such as the DETECT trial [11]. Of particular interest in this study is the large heterogeneity observed in the HER-2 expression of CTCs in breast cancer patients. After evaluation of the HER-2 signal CTCs were divided up into three groups: HER-2 negative, HER-2 low and HER-2 high. This division was based on the HER-2 expression levels measured on MDA-MB 231, MDA-MB 453 and SKBR3 breast cancer cells. A more meaningful division may be to discriminate subgroups based on the effect of a HER-2- targeted treatment

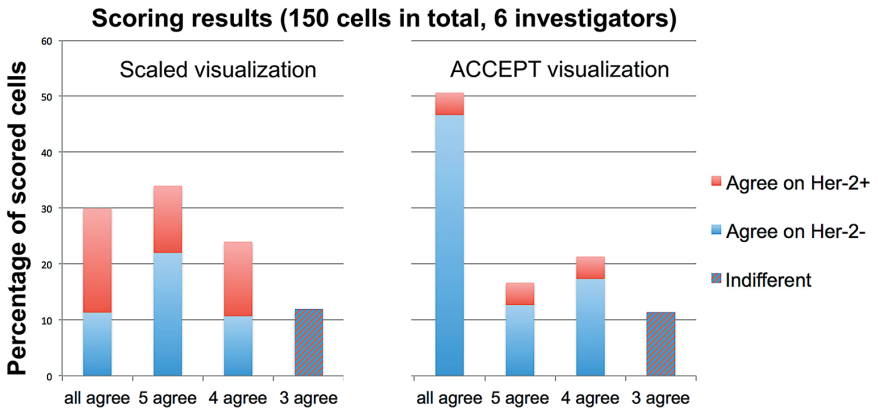
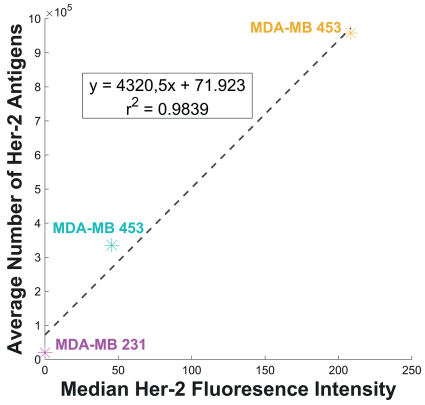
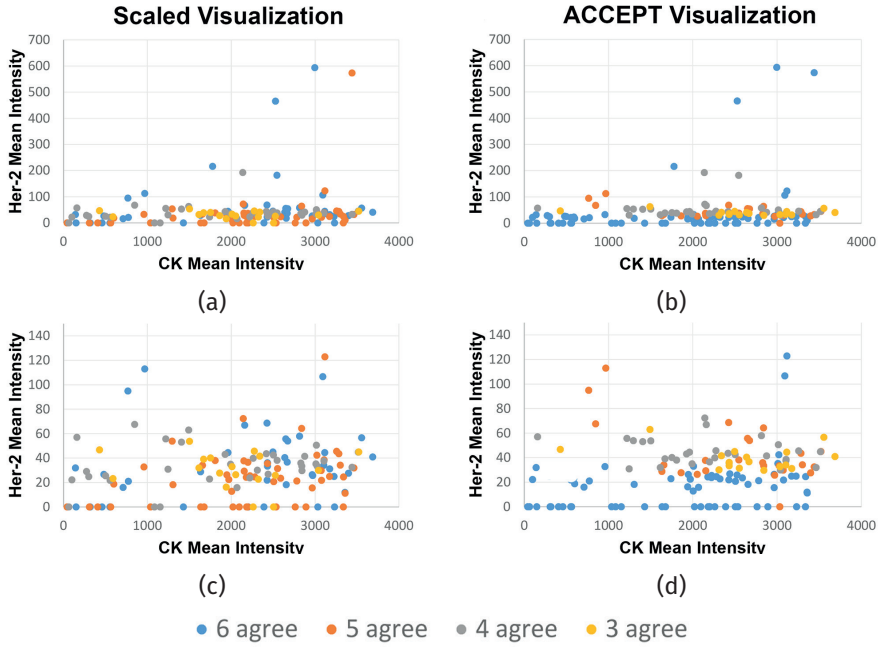


Figure 3.6: *Comparisons of HER-2 assessment by different sites using ACCEPT versus CellTracks Analyzer II ® (Menarini Silicon Biosystems Inc) visualization.* Indicated are the percentages of scored CTCs where all, five, four or only three out of six investigators agreed on the HER-2 status. In the case of 3 agree indifferent means that three investigators vote for HER-2 positive and the other three for HER-2 negative.

on the CTCs directly. This could for example be achieved by measuring HER-2 expression levels of CTCs before and after administration of a HER-2 targeted therapy. In earlier studies, we developed computer generated CTC definitions using overall survival of the patients as the training parameter [14] and extract information from the identified CTCs [14, 15]. In this study, we used advanced mathematical approaches to identify objects in the images and to extract features from the identified objects [25]. More importantly, we make the program available for all researchers with an interest in identifying and characterizing CTCs or other objects and enable the comparison of fluorescence signals generated with different imaging platforms. By using the tools provided in ACCEPT differences in the effectiveness of different treatments targeting the HER-2 receptor can be assessed. The ACCEPT program and manual can be downloaded from https://github.com/LeonieZ/ACCEPT. Besides a more reliable and reproducible quantification of therapeutic marker expression, our ultimate goal is to develop a common definition of a CTC. The large variety of cancer and phenotypes makes it of course very challenging but using the input of a variety of users applying a variety of platforms together with the current breakthroughs in the field of machine learning and imaging may help to understand differences and similarities in the CTCs between different cancer and phenotypes in the future and enables us to find a common CTC definition spanning at least a wide range of cancer and phenotypes. Until then, a unified approach to the analysis of fluorescent images across CTC isolation platforms and CTC phenotypes may increase robustness and lead to a higher reproducibility of results as we have seen it in this study. In its current version, the software can load images from several CTC isolation platforms. Moreover, it contains a general tiff-loader if no specific sample information should be loaded. Loaders for different microscope system can be easily integrated and are work in progress. Further features and tools will be available throughout the IMI CANCER-ID program.



Supplementary Figure 3.1: *Linear calibration of the number of HER-2 antigens and the measured HER-2 signal intensity.* Values plotted for each of the investigated cell lines together with the corresponding line equation and regression value.



Supplementary Figure 3.2: *Comparisons of HER-2 assessment using ACCEPT versus CellTracks Analyzer II ® (Menarini Silicon Biosystems Inc) visualization in relation to measured mean intensities.*Expression of Cytokeratin and HER-2 on the 150 randomly chosen images of CTCs that were sent to six different investigators for scoring HER-2 positivity. Marker colors indicate if all, five, four or only three out of six investigators agreed on the HER-2 status. Panel (a) and (c) correspond to the scaled visualization ((c) is a zoom-in of (a)) and panel (b) and (d) correspond to the ACCEPT visualization ((d) is a zoom-in of (b)).

Bibliography

[1] C. Alix-Panabières and K. Pantel. Challenges in Circulating Tumour Cell Research. *Nature Reviews Genetics*, 14(9):623–631, 2014.

[2] W. J. Allard, J. Matera, M. C. Miller, M. I. Repollet, M. C. Connelly, C. G. Rao, A. G. Tibbe, J. W. Uhr, and L. W. Terstappen. Tumor Cells Circulate in the Peripheral Blood of All Major Carcinomas but not in Healthy Subjects or Patients with Nonmalignant Diseases. *Clinical Cancer Research*, 10(20):6897–6904, 2004.

[3] W. J. Allard and L. W. Terstappen. CCR 20th Anniversary Commentary: Paving the Way for Circulating Tumor Cells. *Clinical Cancer Research*, 21(13):2883–2885, 2015.

[4] G. Attard, J. F. Swennenhuis, D. Olmos, A. H. Reid, E. Vickers, R. A'Hern, R. Levink, F. A. Coumans, J. Moreira, R. Riisnaes, N. B. Oommen, G. Hawche, C. Jameson, E. Thompson, R. C. Sipkema, C. P. Carden, C. Parker, D. Dearnaley, S. B. Kaye, C. S. Cooper, A. Molina, M. E. Cox, L. W. Terstappen, and J. S. de Bono. Characterization of ERG, AR and PTEN Gene Status in Circulating Tumor Cells from Patients with Castration-Resistant Prostate Cancer. *Cancer Research*, 69(7):2912–2918, 2009.

[5] A. M. Barradas and L. W. Terstappen. Towards the Biological Understanding of CTC: Capture Technologies, Definitions and Potential to Create Metastasis. *Cancers*, 5(4):1619–1642, 2013.

[6] M. Benning, C. Brune, M. Burger, and J. Müller. Higher-Order TV Methods - Enhancement via Bregman Iteration. *Journal of Scientific Computing*, 54(2-3):269–310, 2013.

[7] C. Brune, A. Sawatzky, and M. Burger. Primal and Dual Bregman Methods with Application to Optical Nanoscopy. *International Journal of Computer Vision*, 92(2):211–229, 2011.

[8] J. S. de Bono, G. Attard, A. Adjei, M. N. Pollak, P. C. Fong, P. Haluska, L. Roberts, C. Melvin, M. I. Repollet, D. A. Chianese, M. C. Connelly, L. W. Terstappen, and A. Gualberto. Potential Applications for Circulating Tumor Cells Expressing the Insulin-Like Growth Factor-I Receptor. *Clinical Cancer Research*, 13(12):3611–3616, 2007.

[9] D. F. Hayes, T. M. Walker, B. Singh, E. S. Vitetta, J. W. Uhr, S. Gross, C. G. Rao, G. V. Doyle, and L. W. Terstappen. Monitoring Expression of HER-2 on Circulating Epithelial Cells in Patients with Advanced Breast Cancer. *International Journal of Oncology*, 21(5):1111–1117, 2002.

[10] T. J. N. Hiltermann, M. M. Pore, A. van den Berg, W. Timens, H. M. Boezen, J. J. Liesker, J. H. Schouwink, G. J. Wijnands, G. S. Kerner, F. A. Kruyt, H. Tissing, A. G. Tibbe, L. W. Terstappen, and H. J. Groen. Circulating Tumor Cells in Small-Cell Lung Cancer: a Predictive and Prognostic Factor. *Annals of Oncology*, 23(11):2937–2942, 2012.

[11] M. Ignatiadis, S. Riethdorf, F.-C. Bidard, I. Vaucher, M. Khazour, F. Rothé, J. Metallo, G. Rouas, R. E. Payne, R. Coombes, I. Teufel, U. Andergassen, S. Apostolaki, E. Politaki, D. Mavroudis, S. Bessi, M. Pestrin, A. Di Leo, M. Campion, M. Reinholz, E. Perez, M. Piccart, E. Borgen, B. Naume, J. Jimenez, C. Aura, L. Zorzino, M. Cassatella, M. Sandri, B. Mostert,

S. Sleijfer, J. Kraan, W. J. Janni, T. Fehm, B. Rack, L. W. Terstappen, M. I. Repollet, J.-Y. Pierga, M. C. Miller, C. Sotiriou, S. Michiels, and K. Pantel. International Study on Inter-Reader Variability for Circulating Tumor Cells in Breast Cancer. *Breast Cancer Research*, 16(2):R43, 2014.

[12] W. J. Janni, B. Rack, L. W. Terstappen, J.-Y. Pierga, F.-A. Taran, T. Fehm, C. Hall, M. R. de Groot, F.-C. Bidard, T. W. Friedl, P. A. Fasching, S. Y. Brucker, K. Pantel, and A. Lucci. Pooled Analysis of the Prognostic Relevance of Circulating Tumor Cells in Primary Breast Cancer. *Clinical Cancer Research*, 22(10):2583–2593, 2016.

[13] M. G. Krebs, R. Sloane, L. Priest, L. Lancashire, J.-M. Hou, A. Greystoke, T. H. Ward, R. Ferraldeschi, A. Hughes, G. Clack, M. Ranson, C. Dive, and F. H. Blackhall. Evaluation and Prognostic Significance of Circulating Tumor Cells in Patients with Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology*, 29(12):1556–1563, 2011.

[14] S. T. Ligthart, F.-C. Bidard, C. Decraene, T. Bachelot, S. Delaloge, E. Brain, M. Campone, P. Viens, J.-Y. Pierga, and L. W. Terstappen. Unbiased Quantitative Assessment of HER-2 Expression of Circulating Tumor Cells in Patients with Metastatic and Non-Metastatic Breast Cancer. *Annals of Oncology*, 24(5):1231–1238, 2013.

[15] S. T. Ligthart, F. A. Coumans, G. Attard, A. Cassidy, J. S. de Bono, and L. W. Terstappen. Unbiased and Automated Identification of a Circulating Tumour Cell Definition that Associates with Overall Survival. *PLOS ONE*, 6(11):e27419, 2011.

[16] M. Mazel, W. Jacot, K. Pantel, K. Bartkowiak, D. Topart, L. Cayrefourcq, D. Rossille, T. Maudelonde, T. Fest, and C. Alix-Panabières. Frequent Expression of PD-L1 on Circulating Breast Cancer Cells. *Molecular Oncology*, 9(9):1773–1782, 2015.

[17] S. Meng, D. Tripathy, S. Shete, R. Ashfaq, B. Haley, S. Perkins, P. Beitsch, A. Khan, D. Euhus, C. Osborne, E. Frenkel, S. Hoover, M. Leitch, E. Clifford, E. S. Vitetta, L. Morrison, D. Herlyn, L. W. Terstappen, T. Fleming, T. Fehm, T. Tucker, N. Lane, J. Wang, and J. W. Uhr. HER-2 Gene Amplification Can Be Acquired as Breast Cancer Progresses. *Proceedings of the National Academy of Sciences*, 101(25):9393–9398, 2004.

[18] S. Meng, D. Tripathy, S. Shete, R. Ashfaq, H. Saboorian, B. Haley, E. Frenkel, D. Euhus, M. Leitch, C. Osborne, E. Clifford, S. Perkins, P. Beitsch, A. Khan, L. Morrison, D. Herlyn, L. W. Terstappen, N. Lane, J. Wang, and J. W. Uhr. uPAR and HER-2 Gene Status in Individual Breast Cancer Cells from Blood and Tissues. *Proceedings of the National Academy of Sciences*, 103(46):17361–17365, 2006.

[19] M. C. Miller, G. V. Doyle, and L. W. Terstappen. Significance of Circulating Tumor Cells Detected by the CellSearch System in Patients with Metastatic Breast Colorectal and Prostate Cancer. *Journal of Oncology*, 2010:617421, 2010.

[20] J. H. Myung and S. Hong. Microfluidic Devices to Enrich and Isolate Circulating Tumor Cells. *Lab Chip*, 15(24):4500–4511, 2015.

[21] C. G. Rao, D. A. Chianese, G. V. Doyle, M. C. Miller, T. Russell, R. A. Sanders, and L. W. Terstappen. Expression of Epithelial Cell Adhesion Molecule in Carcinoma Cells Present in Blood and Primary and Metastatic Tumors. *International Journal of Oncology*, 27(1):49–57, 2005.

[22] J. S. Ross, J. A. Fletcher, G. P. Linette, J. Stec, E. Clark, M. Ayers, W. F. Symmans, L. Pusztai, and K. J. Bloom. The HER-2/Neu Gene and Protein in Breast Cancer 2003: Biomarker and Target of Therapy. *The oncologist*, 8(4):307–325, 2003.

[23] J. B. Smerage, G. T. Budd, G. V. Doyle, M. Brown, C. Paoletti, M. Muniz, M. C. Miller, M. I. Repollet, D. A. Chianese, M. C. Connelly, L. W. Terstappen, and D. F. Hayes. Monitoring Apoptosis and Bcl-2 on Circulating Tumor Cells in Patients with Metastatic Breast Cancer. *Molecular Oncology*, 7(3):680–692, 2013.

[24] G. van Dalum, G.-J. Stam, L. F. Scholten, W. J. Mastboom, I. Vermes, A. G. Tibbe, M. R. de Groot, and L. W. Terstappen. Importance of Circulating Tumor Cells in Newly Diagnosed Colorectal Cancer. *International Journal of Oncology*, 46(3):1361–1368, 2015.

[25] L. L. Zeune, G. van Dalum, L. W. Terstappen, S. A. van Gils, and C. Brune. Multiscale Segmentation via Bregman Distances and Nonlinear Spectral Analysis. *SIAM Journal on Imaging Sciences*, 10(1):111–146, 2017.

Chapter 4

Classification of Cells in CTC-Enriched Samples by Advanced Image Analysis

Sanne de Wit*, Leonie L. Zeune*, Jeroen N. Hiltermann,
Harry J.M. Groen, Guus van Dalum and Leon W.M.M. Terstappen

Cancers 10.10 (2018): 377

* Both authors contributed equally.

Abstract

In the CellSearch® system, blood is immunomagnetically enriched for epithelial cell adhesion molecule (EpCAM) expression and cells are stained with the nucleic acid dye 4'6-diamidino-2-phenylindole (DAPI), Cytokeratin-PE (CK), and CD45-APC. Only DAPI+/CK+ objects are presented to the operator to identify circulating tumor cells (CTC) and the identity of all other cells and potential undetected CTCs remains unrevealed. Here, we used the open source imaging program Automatic CTC Classification, Enumeration and PhenoTyping (ACCEPT) to analyze all DAPI+ nuclei in EpCAM-enriched blood samples obtained from 192 metastatic non-small cell lung cancer (NSCLC) patients and 162 controls. Significantly larger numbers of nuclei were detected in 300 patient samples with an average and standard deviation of 73,570 ± 74,948, as compared to 359 control samples with an average and standard deviation of 4191 ± 4463 ($p < 0.001$). In patients, only 18% ± 21% and in controls 23% ± 15% of the nuclei were identified as leukocytes or CTC. Adding CD16-PerCP for granulocyte staining, the use of an LED as the light source for CD45-APC excitation and plasma membrane

staining obtained with wheat germ agglutinin significantly improved the classification of EpCAM-enriched cells, resulting in the identification of 94% ± 5% of the cells. However, especially in patients, the origin of the unidentified cells remains unknown. Further studies are needed to determine if undetected EpCAM+/DAPI+/CK–/CD45– CTCs are present among these cells.

4.1 Introduction

Circulating tumor cells (CTC) are cells disseminated from the primary or metastatic tumor site into the bloodstream. In several cancer types, including non-small and small cell lung, prostate, breast, colon, bladder, gastric carcinoma, and melanoma, the presence of CTCs in circulation is associated with a poor survival rate of the patients [8, 10, 11, 15, 17, 18, 22, 25, 26]. With many blood cells circulating in our cardiovascular system, CTCs are rare events. A concentration of one CTC per mL of blood is already high, while the concentration of white blood cells is around 5×10^6 cells and red blood cell concentrations reach up to 5×10^9 cells per mL. When detecting CTCs, it is therefore of utmost importance to locate all CTCs present and have an extremely sensitive and specific assay to discriminate between the tumor cells and blood cells. In the CellSearch® system, epithelial cell adhesion molecule (EpCAM) antibodies coupled to ferrofluids are used for immunomagnetic enrichment of cells thereby removing the bulk of leukocytes. After this, CTCs are fluorescently labeled with: 4'6-diamidino-2-phenylindole (DAPI) to stain the nucleus; phycoerythrin (PE) labeled antibodies recognizing cytokeratins (CK) 4–6, 8, 10, 13, 18, and 19, which are present in the majority of cells derived from epithelial cancers, and allophycocyanin (APC) labeled cluster of differentiation (CD) 45, which is present on the cell surface of leukocytes [2]. Images covering the surface of the CellTracks cartridge are acquired by the CellTracks Analyzer II, and all events that are positive for both DAPI and CK are presented as thumbnail images to the operator, who will decide which cells follow the criteria for CTC assignment. The identity and number of all other nucleated cells are not provided, thereby obscuring any potential cells of interest. These cells can now be explored using the open source image analysis program Automatic CTC Classification, Enumeration and PhenoTyping (ACCEPT) (download on https://github.com/LeonieZ/ACCEPT). The ACCEPT program is an open source toolbox designed to analyze and characterize all objects present in the fluorescent images. With this program, we analyzed 659 EpCAM-enriched blood samples, obtained from 162 healthy donors or patients with benign disease and 192 metastatic non-small cell lung cancer (NSCLC) patients. We pursued several methods to determine the extent and identity of all cell populations present, in order to assess the number of cells that potentially are CTCs. We observed that the number of cells identified as leukocytes in the EpCAM-enriched fraction was low and showed a large variance across the samples. We investigated several possible causes: A possible correlation with sample or patient attributes, insufficiencies in the image analysis, and insufficiencies in the staining of the leukocytes or its detection. To improve this last option, we explored several avenues. We showed previously that addition of CD16, recognizing the FcγIIIa, to the staining cocktail improved the identification of nu-

clei as granulocytes since they express CD45 at a lower level compared to lymphocytes and monocytes [33]. We now applied this antibody in a large cohort of patient samples. We also explored an alternative LED excitation light source to improve the excitation efficiency of the APC fluorochrome, as opposed to the mercury arc lamp which is used as a light source in the CellTracks [24]. In addition, wheat germ agglutinin (wga), a lectin that binds to sialic acid and N-acetylglucosamine residues, was explored because it is present on cellular plasma membranes [1, 5, 13, 23, 30, 31]. We examined the presence of a cell membrane by using wga conjugated to AlexaFluor-488 and adding it to the CellSearch immunostaining cocktail. In summary, we present a method to closely examine every cell present in the cartridge after EpCAM enrichment. CTC subpopulations expressing low or no CK are currently not examined, and their potential clinical utility remains unexploited. Here we describe several methods to discover the extent of the anonymous cell population that could harbor these unidentified CTCs.

4.2 Materials and Methods

4.2.1 Cancer Patients, Patients with Benign Disease, and Healthy Volunteers

Peripheral blood samples were drawn by venipuncture into 10 mL CellSave Preservative Tubes (Menarini Silicon Biosystems, Huntingdon Valley PA, USA) from healthy volunteers and metastatic NSCLC patients. A total of 300 blood samples from 192 NSCLC patients were obtained after patients with advanced NSCLC provided written informed consent. The medical ethical committee of the University Medical Center Groningen and Board of Directors approved the protocol (project number 288.194; April 7th, 2014; Groningen, The Netherlands). Blood draw from patients occurred before and/or during their treatment. A total of 127 blood samples from 20 healthy volunteers aged 20–55 were obtained from the donor services of the Experimental Centre for Technical Medicine at the University of Twente, and all donors gave written informed consent before donating blood. All healthy volunteer samples were processed 1 day after blood draw, whereas patient samples were processed after 1–4 days. Digitally stored images from 232 samples from 142 patients with benign disease processed with CellSearch from previous studies were included for ACCEPT analysis [2, 32].

4.2.2 Cell Lines and Spiking

Whole blood samples from healthy volunteers were spiked with cancer cell lines PC3, NCI-H460, and MCF-7. PC3 (1.0×104 EpCAM antigens) and NCI-H460 (1.4×102 EpCAM antigens) were used for frequency detection of CTCs with ACCEPT and MCF-7 (8.8×105 EpCAM antigens) was used for testing wheat germ agglutinin (wga) staining of cancer cells. This cell line was chosen for its high EpCAM expression to ensure sufficient capture of cancer cells in the cartridge for analysis. All cell lines were obtained from ATCC (Manassas, VA, USA) and have not been authenticated in the past six months. The cells were grown at 37°C and 5% CO2 and

cultured in RPMI-1640 (Sigma-Aldrich, St. Louis MO, USA). The culture media was supplemented with L-Glutamine (Sigma), 10% fetal bovine serum (Gibco, Invitrogen, Carlsbad CA, USA), and 1% penicillin-streptomycin (Gibco, Invitrogen, Carlsbad CA, USA).

4.2.3 Processing Blood with CellSearch

Aliquots of 7.5 mL of blood were processed with the CellSearch® system (Menarini Silicon Biosystems, Castel Maggiore (BO), Italy) for CTC detection. CellSearch analysis was performed within 96 h after blood was drawn. Antibodies directed against the epithelial cell adhesion antigen (EpCAM) coupled to ferrofluids were used to enrich CTCs from the background of leukocytes. The enriched cells were fluorescently labeled with the CellSearch CTC kit (Menarini Silicon Biosystems) using the nucleic acid dye 4'6-diaminodino-2-phenylindole (DAPI) for DNA staining, anti-cytokeratin monoclonal antibodies (mAbs) C11 and A53.B/A2 labelled with phycoerythrin (PE), and anti-CD45 mAb (clone HI30) labeled with allophycocyanin (APC) for recognizing leukocytes. Peridinin Chlorofyll A Protein (PerCP) labeled mAb anti-CD16 (clone 3G8, Biolegend, San Diego, CA, USA) directed against granulocytes, and Alexa-Fluor488 conjugated to the lectin wga (Thermo Fisher Scientific, Waltham, MA, USA) were added to the extra marker position in the CellSearch Epithelial Cell kit with an end concentration of 2 μg/mL and 3 μg/mL, respectively. This extra marker channel was used for the measurement of PerCP and DIOC on the CellTracks Analyzer II, which generates images of the complete cartridge for all channels.

4.2.4 Image Acquisition

fluorescence microscopy images were acquired using a modified CellTracks Analyzer II, which included a 20X and 40X objective and in total six filter cubes next to the standard instruments. Besides the acquisition of the standard channels DAPI, PE, DIOC, and APC, the images with PerCP staining were acquired with a customized PerCP filter cube (excitation 435/40 nm; dichroic 510 nm; emission 676/29 nm (Semrock, Rochester, NY, USA)) using an integration time of 400 ms. When wga was used in the immunostaining, the integration time of the DIOC filter was also 400 ms. To determine if another light source detects weakly stained leukocytes better than the mercury arc lamp of the CellTracks, ten cartridges were automatically scanned with an inverted Nikon Ti-E microscope with filter cube and objective changer (Nikon, Tokyo, Japan). The light source of this microscope is the Lumencor Sola SE II 365 LED. A 10X (0.45NA) objective (Nikon, Tokyo, Japan), a computer-controlled CCD camera (C114400, Orca Flash 4.0LT, Hamamatsu Photonics, Hamamatsu, Japan), an ASI MS-2500 XY stage and filters optimized for LED illumination were used for the acquisition of fluorescent images of the cartridge. The following filters (all from Semrock) were used: DAPI with excitation 377/50 nm, dichroic 409 nm, emission 409 nm/LP; fluorescein isothiocyanate (FITC) with excitation 482/18 nm, dichroic 495 nm, emission 520/28 nm; PE with excitation 543/22 nm, dichroic 562 nm, emission 593/40 nm; APC with excitation 635/18 nm, dichroic 652 nm, emission 680/42 nm; and PerCP with excitation 435/40 nm, dichroic 510 nm, emission 676/29 nm. The integration times were the same as used in the CellTracks: For DAPI 30 ms, FITC 100 ms, PE 200

ms, APC 600 ms, and for PerCP 400 ms. The scanning and image acquisition was controlled by a software program written in Labview (National Instruments, Austin, TX, USA).

4.2.5 Image Analysis with ACCEPT

In CellSearch, CTCs are annotated when the objects are stained with DAPI and CK, but lack CD45 staining and are larger than 4 μm with morphological features consistent with those of a cell [2]. In the ACCEPT (Automated CTC Classification, Enumeration, and PhenoTyping) toolbox, every event in the images is analyzed using an advanced multi-scale segmentation approach and several intensity and shape measurements are extracted for every event found [38, 39]. ACCEPT is developed in the EU funded "CANCER-ID" and "CTC-Trap" programs, as an open source toolbox for CTC analysis and can be downloaded from *https://github.com/LeonieZ/ACCEPT*. New tools and features for ACCEPT are being developed and made available in the downloadable version at regular intervals. For this study, the fluorescent images from all samples were uploaded, and all events present in the images were detected. To enhance the detection of cells, especially in areas where the cells formed large cell clusters, we added a background subtraction to the ACCEPT toolbox. Before images are segmented, a Fourier filtering is applied, and large and very fine scales are filtered by multiplying the filtered signal with two Gaussian kernels before the signal is back-transformed by an inverse Fourier transformation. In this way, the background is removed, which enhances the detection in very crowded areas, as well as areas with inhomogeneous illumination. Fragmented fluorescent signals make a segmentation with an intensity-based segmentation method difficult, which is now implemented in the downloadable version of the ACCEPT toolbox. In this study this was especially observed in the fluorescent images obtained from the CD16-PerCP stained cells. Therefore, we decided to explore deep learning approaches to determine if this limitation could be overcome. For this approach, all thumbnails containing nucleated cells in the cartridge were found using ACCEPT, and a random selection of these thumbnails was used for further evaluation. Per sample, we exported 10% of the observed cells, but minimally 1000 (if the total amount was high enough) and maximally 5000 thumbnails. Moreover, we exported 759 thumbnails with five fluorescent channels each where the segmentation method of ACCEPT gave satisfactory results. We used these thumbnails (split into single channels) to train a deep learning network to segment the cells. The implementation was done in Keras and we used the U-net architecture with skip connections [7, 27]. The set of 3796 thumbnail images was split into a training set of 3037 images (80%) and a validation set of 759 images (20%). The images were preprocessed in the following way: Each thumbnail exported from the ACCEPT toolbox was transformed into a thumbnail of 80×80 pixels by adding a padding at all four sides with the median intensity of the pixels at the image boundary. Moreover, all images scanned with CellTracks were cast from double to uint8 arrays by first dividing by 4095 (the maximum value in Celltracks) and then multiplying by 255 and converting the results to uint8. Images scanned with the inverted Nikon Ti-E microscope were recorded as uint16, thus having a theoretical maximum value of 65,535. However, during our analysis, we noticed that this range was not used and that the range of intensity values used strongly differs between the fluorescent channels. Therefore, we

calculated for each fluorescent channel the intensity value that was higher than 99.9% of the measured intensities over all images in all samples. The images were then divided by the resulting intensity per channel, multiplied by 255 and cast to uint8. By taking the value that is above 99.9% instead of the maximum, we accounted for individual saturated pixels where the incident light caused the camera sensor to respond with the highest possible value. To train the deep learning network, we used a batch size of 16 and the network was trained for ten epochs. The resulting convolutional network was then used to segment the exported thumbnails for each sample. The original thumbnails together with the segmentation derived by deep learning were fed back into the ACCEPT algorithm by using the marker characterization processor for further analysis and visualization of results.

4.2.6 Statistical Analysis

Statistical analysis was performed using SPSS (version 24, SPPS Inc., Chicago, IL, USA). A p-value below 0.05 was considered significant. The non-parametric Mann-Whitney U test with significance level $\alpha = 0.05$ was used to determine if there was a difference in nucleated cell numbers between cohorts of patients and controls. The total nucleated cell amounts were log-transformed for normality. With multiple regression analysis, we determined the association of several variables with the total nucleated cell count, such as age, treatment, CTC score, and sample age before processing. The confounding effect of each variable was also determined.

4.3 Results

4.3.1 Enumeration of Nucleated Cells in the EpCAM-Enriched Cells

The image analysis program ACCEPT was used to enumerate the total number of nucleated cells in the image files after EpCAM enrichment. Figure 4.1 (a) – (c) shows an example of the ACCEPT analysis in which the nucleated cells (blue dots) are identified among the other objects in the images (grey dots). To identify nucleated cells in the cartridge images, showing the following criteria for DNA staining were used: 1. Mean fluorescence intensity ≥ 50 (panel (a)); 2. standard deviation of the fluorescence intensity ≥ 20; 3. Size $< 500\ \mu m^2$ (panel (a)); and 4. Roundness < 0.95, where 0 is a perfect circle (panel (b)). To identify single cells, a perimeter to area (P2A) of < 1.5 was used, for doublets a P2A between ≥ 1.5 and < 2.5, for small clusters a P2A between ≥ 2.5 and < 4, and for large clusters a P2A ≥ 4 (panel (b)). For enumeration of nucleated cells, the doublets and clusters were counted as one. An example of the single cells, double cells, and clusters that were identified in this manner are illustrated in Figure 4.1 (d) – (g). Figure 4.2 shows the nucleated cell count from 300 samples obtained from 192 NSCLC patients (mean 73,570 ± 74,948, median 47,192) and from 359 samples obtained from 162 healthy donors and patients with benign disease (mean 4,191 ± 4463, median 2729). The number of nucleated cells was significantly larger in the patient samples as compared to the samples from healthy donors and patients with

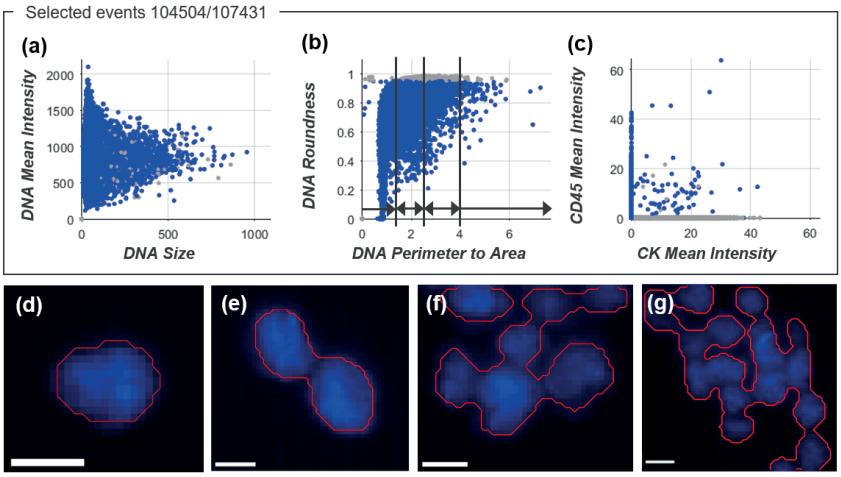


Figure 4.1: *ACCEPT identification of nucleated cells.* The three scatter plots in panels (a) – (c) were used to define nucleated cells (depicted in blue) by ACCEPT. In total, 104,504 events were detected and 107,431 of them were classified as nucleated cells, whereas the other events are depicted as grey dots. The division into single cells, doublets, small, and large clusters was based on DNA perimeter to area ratio, as illustrated in panel (b). In the scatter plot in panel (c), the mean fluorescence intensity of Cytokeratin (CK)-phycoerythrin (PE) (CK mean intensity) is plotted against the mean fluorescence intensity of CD45-allophycocyanin (APC) (CD45 mean intensity). In panels (d) – (g) typical examples of the segmentation (red line) around the nucleus of a single cell are illustrated in panel (d); of a doublet in panel (e); a small cluster in panel (f); and a large cluster in panel (g). The white scale bar represents 10 pixels, which corresponds to a size of 6.4 μm.

benign disease (Mann-Whitney U test: $p < 0.001$). For a complete analysis, we determined the association between the number of nucleated cells and the sample type ("patients" or "controls"). For this analysis, 300 samples from metastatic NSCLC patients and 359 samples from healthy volunteers and patients with benign disease were used. First, several factors were investigated that could influence the nucleated cell count since these factors were present in a limited set of patients or controls only. As the first factor, we evaluated the influence of the assay itself by determining the nucleated cell count from 30 controls with benign disease from which four blood tubes were obtained and processed simultaneously, see Supplementary Figure 4.1. It indicated that a coefficient of variation of less than 10.7% or more than 40.2% (mean of 25.6% ± 14.9% (1 standard deviation)) of the nucleated cell count cannot be attributed to assay variation with a 97% certainty. This was only the case for six (20%) controls. The influence of the assay itself was therefore excluded from further analysis.

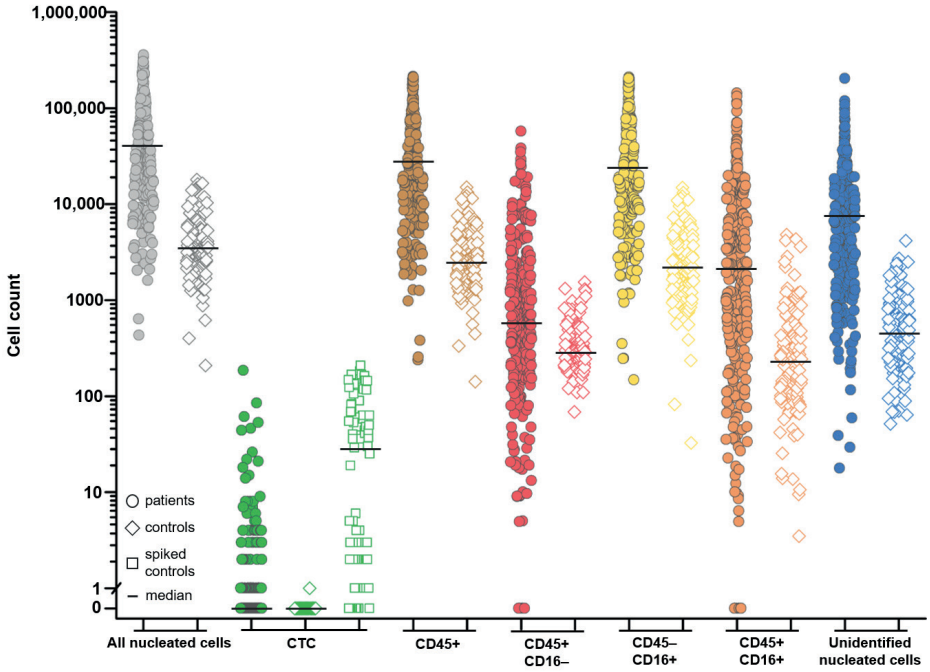


Figure 4.2: *Frequency distribution of the cell populations detected in CellSearch® and with the addition of the CD16 classification after epithelial cell adhesion molecule (EpCAM) immunomagnetic enrichment of 300 blood samples from 192 non-small cell lung cancer (NSCLC) patients (filled circle) and 127 blood samples from 20 healthy volunteers (open diamond).* Healthy volunteer samples spiked with cell line cells (n = 88) are indicated with an open square. Cell populations: All nucleated cells (grey), circulating tumor cells (CTC) (green), CD45+ leukocytes (brown), CD45+/CD16– leukocytes (red), CD45–/CD16+ leukocytes (yellow), CD45+/CD16+ leukocytes (orange), and unidentified cells (blue). The differences in cell count between the NSCLC patients and healthy volunteers in the same cell classification is significant for all subclasses ($p < 0.001$). Samples from healthy volunteers were not spiked (diamond symbol, n = 39; median 0 CTC) or spiked (square symbol) with cancer cell lines PC3 (n = 47; median 63 CTC; 1.0×104 EpCAM antigens) and NCI-H460 (n = 41; median 1 CTC; 1.4×102 EpCAM antigens).

Second, undergoing treatment as a contributing factor was considered. In 64 NSCLC patients, we observed that 95% of the patients showed an increase or decrease in nucleated events of more than 10% during treatment, compared to the nucleated events at the start of treatment, see Supplementary Figure 4.1. We considered this a significant change and therefore this factor was included in further analysis. The third factor investigated was gender. No influence of the gender (133 male patients, 127 female patients, and 40 patients of unknown gender) on the nucleated cell count was observed with the Mann-Whitney U test ($p = 0.237$), and the factor was therefore excluded from further analysis. In the crude multiple regression analysis for all 300 NSCLC patients and 359 controls, the relation between

	Regression coefficient	95% Confidence Interval	*p*-value
Crude analysis	17.0	14.1 – 20.6	< 0.001
Corrected analysis for *age*	19.2	15.3 – 24.2	< 0.001
Corrected analysis for *age* and patient variables (*sample age*, *CTC count* and *treatment*)	12.4	9.4 – 16.4	< 0.001

Table 4.1: The definition of cell populations used to classify cells in ACCEPT after deep learning segmentation in 300 non-small cell lung cancer (NSCLC) patient samples and 127 samples from healthy volunteers.

sample types “patient” and “control” was established. The result of this multiple regression analysis is presented in Table 4.1. This was followed by an analysis to correct for patient age (younger vs. older than 55 years). As a third analysis, we corrected for the patient variables: CTC count (continuous variable), treatment (during treatment vs. before or no treatment) and sample age at the time of processing with CellSearch (1 day vs. 2–4 days). Each variable was analyzed for its confounding effect on the number of nucleated events. This showed that age of the patient, undergoing treatment, and CTC count had little confounding effect (below 5%). The sample age showed a confounding effect of 11.7%.

4.3.2 Improved Image Analysis

The number of nucleated events present in some cartridges can be enormous (over 100,000 for 27% of the patients), covering the whole detection surface, while some cartridges barely hold any cells (less than 5000 for 6% of the patients). The high number of cells creates a large area of packed cells, which makes it difficult to define borders for each cell, either manually or by image analysis algorithms. The segmentation quality of the ACCEPT toolbox was influenced by the number of cells present in the sample, an illumination bias, and fragmented staining. This is illustrated in detail in Supplementary Figure 4.3. These problems motivated the use of a segmentation approach based on deep learning. As opposed to the model-based approaches currently implemented in the ACCEPT toolbox, deep learning approaches do not make any model assumptions about the signal (e.g., a nearly homogenous background and cell intensity), but learn a model based on only the data used for the training.

4.3.3 Assignment of Nucleated Cells to a Cell Lineage

In the CellSearch system, CD45-APC is used to identify leukocytes and Cytokeratin-PE to identify CTCs among the nucleated cells. Nucleated cells expressing both CD45-APC and Cytokeratin-PE are assigned as cells binding non-specifically to at least one of the antibodies. The number of CTCs identified by the classical CellSearch method in the 300 samples from 192 NSCLC patients ranged from 0–186 (mean 2; median 0) and represent only a very

		Mean Intensity value					Overlay nucleus			Size
							with	*with*	*with*	(μm^2)
		DAPI	*CD45*	*CK*	*CD16*[1]	*wga*[1]	*CK*	*CD45*	*CD16*	*CK*
CTC	*CK+/CD45–*	> 0	0	> 50	0	> 0	> 0.4	ND	ND	> 9
	CD45+/CD16–	> 0	> 0	0	0	> 0	ND	> 0	ND	ND
Leukocytes	*CD45+/CD16+*	> 0	> 0	0	> 0	> 0	ND	> 0	> 0	ND
	CD45–/CD16+	> 0	0	0	> 0	> 0	ND	ND	> 0	ND
Nuclei	*Bare nucleus*	> 0	0	0	0	0	ND	ND	ND	ND
	Unstained cell	> 0	0	0	0	> 0	ND	ND	ND	ND

Table 4.2: The definition of cell populations used to classify cells in ACCEPT after Deep Learning segmentation in 300 NSCLC patient samples and 127 control samples.

small portion of the nucleated cells identified in the samples from these patients. To determine whether the identity of the majority of nucleated cells could be assigned to the hematopoietic cell lineage by their expression of CD45, the thumbnails of the nucleated cells were put into a trained deep learning network for segmentation and the segmented images were fed back into the ACCEPT toolbox. Using the marker characterization processor tool, the results could be visualized and analyzed. The improved segmentation allowed for a relatively simple definition of cell subpopulations and in most cases a simple yes or no could be used to define the presence or absence of a marker. The actual definitions of the different cell populations are listed in Table 4.2. In this manner, nucleated cells were classified as CD45+, CD45– and CK– or CK+ in 300 samples from 192 NSCLC patients and 127 samples from 20 healthy volunteers.

As described in the method section, for this analysis, 10% of the nucleated cells found by the ACCEPT toolbox, up to a maximum number of 5000 cells, were analyzed and the results were extrapolated to the full number of nucleated events to arrive at the cell numbers. The number of nucleated cells that could be identified as leukocytes in CellSearch was only 18% ± 21% (range 0–89%; median 9%) in patients and 23% ± 15% (range 2–68%; median 19%) in healthy volunteers, see Figure 4.3. To be able to assign the cell lineage of origin of the nucleated cells on which no CD45- or CK-expression could be detected, several approaches were pursued.

4.3.4 Increasing Leukocyte Identification by Adding CD16 Immunostaining

Whereas CD45 is basically present on all human leukocytes, it is most strongly expressed on lymphocytes and monocytes and expressed at a relatively low density in granulocytes. CD16 identifies the FcγIIIa receptor present on granulocytes and natural killers cells. We added anti-CD16 labeled to PerCP to the CellSearch staining cocktail to improve the identification of granulocytes especially, using 300 blood samples from 192 NSCLC patients and 127 blood samples from 20 healthy volunteers. Thumbnail images of the nucleated cells identified

[1]Only for those experiments in which CD16 or wga was added to the immunostaining cocktail

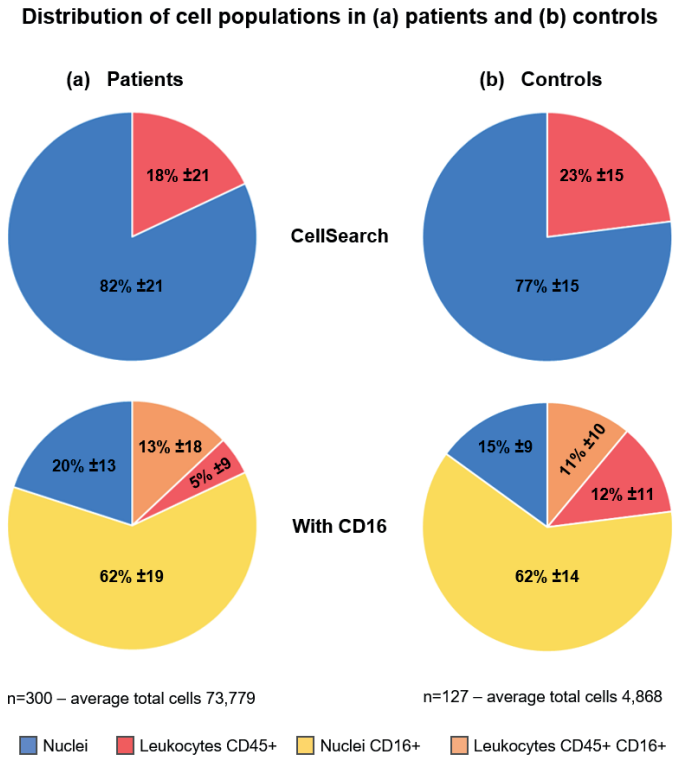


Figure 4.3: *Cell population distribution in (a) patients and (b) controls using the CellSearch definition (upper) and using CD16 expression (lower) for further classification of the cells.*

by ACCEPT were fed into the trained deep learning network as described above. Figure 4.4 shows typical thumbnail images of nucleated events identified by ACCEPT using the presence of the different markers for classification. Classified CTCs with expression of CD45 or CD16 can either be; (1) CTCs nonspecifically binding to CD16 or CD45 antibodies; or (2) leukocytes non-specifically binding to the cytokeratin antibodies. The distribution of these cell populations in the patient and healthy volunteer samples is shown in Figure 4.3. From the 82% ± 21% unidentified nucleated cells in 192 NSCLC patients, 62% ± 19% are identified as granulocytes through CD16-expression. This decreases the number of unidentified cells to 20% ± 13% of the total nucleated cells present in the cartridge. The percentage of CD45-expressing leukocytes that also express CD16 is 13% ± 18%. These are most likely natural killer cells that express CD45 in higher densities compared to granulocytes. In 127 samples from 20 healthy volunteers, 77% ± 15% of the cells are unidentified which was decreased to 15% ± 9% when CD16 was used. The actual number of nucleated cells identified by either CD45 or CD16 for patients and healthy volunteers is shown in Figure 4.3. The number of all classifications of nucleated cells was significantly larger in patients compared to healthy volunteers ($p < 0.001$). Figures 4.2 and 4.3 show there is still a large number of nucleated cells, especially in the patients, that remain unidentified.

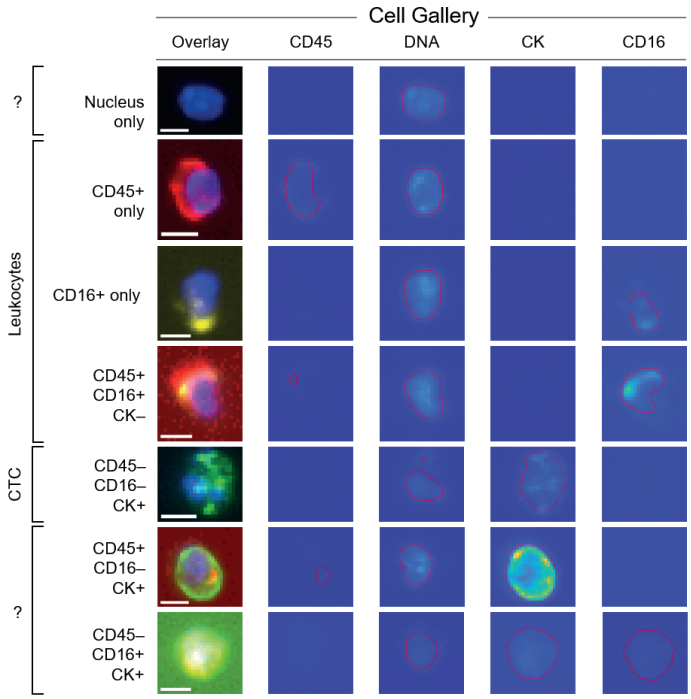


Figure 4.4: *ACCEPT gallery showing cells from all classifications using the presence or absence of several markers.* The scale bar is 10 pixels, representing 6.4 μm.

4.3.5 LED as Light Source to Improve Excitation Efficiency of CD45-APC and CD16-PerCP

For five NSCLC patients and five healthy volunteers, cartridges were scanned first with the CellTracks – equipped with a mercury arc lamp as a light source – followed by scanning on a microscope equipped with an LED light source. The objectives and filter cubes were kept the same in both systems. The mean percentage of CD45+/CD16– leukocytes found by the CellTracks was 8% ± 7% and this increased to 16% ± 9% when a LED light source was used, as shown in Figure 4.5. More pronounced was the increase for CD45+/CD16+ leukocytes, which increased from 1% ± 0% to 59% ± 25% with the LED light source. Whereas CD16+/CD45– cells comprise a large population of the cells when a mercury arc lamp is used, these cells also appear to be CD45+ when the LED light source was used for the analysis; CD16+/CD45– leukocytes decreased from 58 ± 17% to 19% ± 15%. In total, the number of unidentified cells was reduced from 33% ± 14% to 6% ± 5% when an LED light source was used instead of a mercury arc lamp.

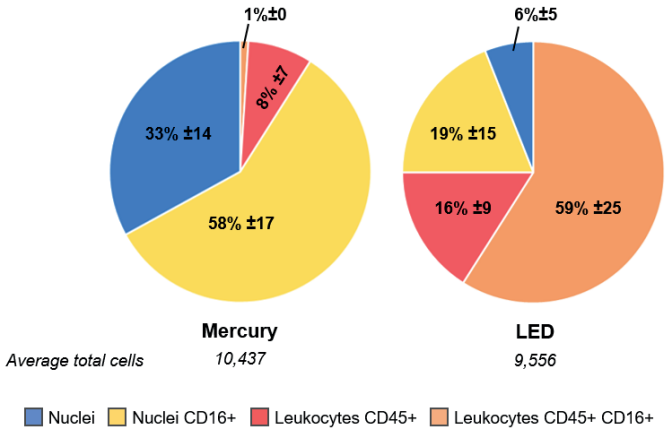


Figure 4.5: *Improved identification of cells by comparison of cell population distributions analyzed with the mercury arc light source in CellSearch (left), followed by analysis with a LED light source on a separate microscope (right).*

4.3.6 Identification of Unstained Nuclei by Adding Wheat Germ Agglutinin Immunostaining

In ten whole blood samples from five NSCLC patients and five healthy volunteers, the lectin wga conjugated to Alexa-Fluor488 was added to the CellSearch staining cocktail for the identification of cellular plasma membranes. In Supplementary Figure 4.2, an ACCEPT gallery shows the staining of wga present on all cell classifications. The absence of wga on a nucleus, classifies this event as a "bare nucleus". The presence of wga on already identified cells was 44% ± 23% for CD16+ granulocytes, 41% ± 20% for CD45+ leukocytes and 88% ± 15% for CD45+/CD16+ leukocytes, see Figure 4.6. One healthy volunteer sample was spiked with the cancer cell line MCF-7, and 100% of those spiked cells were stained with wga. In contrast, 52% ± 23% of the unidentified cell population was stained with wga, suggesting these are intact cells, whereas the remaining 48% ± 23% of these nuclei appear to be without a cell membrane. This suggests that these are simply bare nuclei, which cannot be identified through immunofluorescence staining.

4.4 Discussion

In this study, we investigated the cell populations present in 659 cartridges after immunomagnetic EpCAM enrichment by the CellSearch system, gathered from 162 healthy volunteers and 192 NSCLC patients. Stored fluorescent microscope images of the surface of the cartridges were analyzed by image analysis for the presence of different cell populations. The identity and number of cells present in the EpCAM-enriched cell suspension is not known, since the software only detects and presents DAPI+/CK+ events for review that are co-located. Any potential tumor cells or other interesting events remain obscured. We therefore

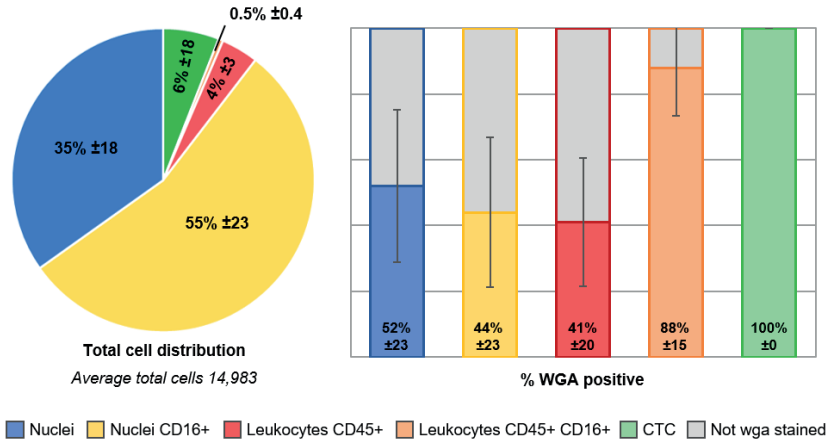


Figure 4.6: *Distribution of cell populations in samples with wheat germ agglutinin (wga)-AlexaFluor488 staining (n = 10).* The stacked bar represents the five cell populations present in these cartridges. One healthy volunteer sample was spiked with MCF-7 cancer cells, which are represented in the "CTC" category (6%). Of each population, the percentage of cells that stained positive for wga is displayed on the right side of the image, whereas the population remaining unstained with wga is visualized in grey.

explored the identity of all events present in the cartridge after EpCAM enrichment. Images from the CellTracks microscope were analyzed with advanced imaging techniques. Most of the image analysis tools that were used are already assembled in the open source imaging program ACCEPT, including a tool to compare operator variability in scoring CTCs [37, 39]. In this image data set we had to deal with samples of low image quality, unequal distribution of the illumination, and extremely crowded images containing many clustered events, see Supplementary Figure 4.3. A deep learning based approach appeared to be a good solution since it could easily be adapted to a certain type of signal to segment by feeding it with enough training data. The deep learning based segmentation algorithm was programmed in Python® using the Keras [7] framework and is not yet linked to the ACCEPT toolbox, which is developed in MATLAB®. To be able to use the deep learning based segmentation, we exported thumbnails identified by ACCEPT and used them as input for the deep learning segmentation network. After segmentation of the thumbnails by the network, the segmentation results, together with the original thumbnails, were fed back into the ACCEPT toolbox for further analysis. For samples with thousands of nucleated events (up to 300,000 with five fluorescent channels each), this became quite a computationally intensive task. Therefore, for the current analysis we restricted that part of our analysis to only 10% of the nucleated events which were randomly selected. Experience obtained with the deep learning based classification algorithm we developed showed very promising results and the new segmentation method also showed that it is often superior to the more traditional segmentation approach currently used in ACCEPT [36]. Thus, in the future, we would like to incorporate a semantic deep learning segmentation into the ACCEPT toolbox to automatically detect and

classify CTC, tumor derived extra cellular vesicles, leukocytes, and other cell populations for which reagents are added. In total, 300 samples obtained from 192 metastatic lung cancer patients, 232 samples obtained from 142 patients with benign tumors, and 127 samples obtained from 20 healthy volunteers were investigated for the presence of nucleated events. The number of nucleated events was significantly higher in cancer patients as compared to the healthy donors and patients with benign disease, see Figure 4.2 and Supplementary Figure 4.1, and further studies are needed to explore the potential clinical utility of this finding. CTCs were detected by traditional CellSearch scoring in a small portion of the patient samples and in these samples the CTC number ranged from 1 to 186 cells. This low number of cells could not account for the large differences in the detected nucleated events. Knowledge of the composition of these nucleated cells is needed to be able to hypothesize about this difference. Most likely, these nucleated events originate from hematopoietic cells that are captured along with the EpCAM enrichment in CellSearch. However, the fraction of nucleated cells on which the leukocyte marker CD45 was detected was remarkably low both in healthy volunteers and metastatic cancer patients, as shown in Figure 4.3. Since the excitation of CD45-APC with a mercury arc lamp used in the CellTracks system is not optimal, we first looked into a way to increase the detection of granulocytes especially, which express the CD45 antigen in relatively low antigen density compared to leukocytes. We evaluated this by adding anti-CD16-PerCP antibody to the immunostaining in the CellSearch system. All 300 samples from metastatic lung cancer patients and 127 samples from healthy volunteers were processed with this extra marker. The leukocytes and granulocytes that were identified indeed increased, as shown in Figure 4.3. However, a considerable number of nucleated cells remained unidentified. Next, we looked into the role of ferrofluids, which are used at a concentration of 40 μg/mL in CellSearch and are present on the imaging surface of the cartridge [9, 28]. The ferrofluids reduce the intensity of the fluorescence; an effect which is most obvious in the blue region (400–460 nm). Therefore, we looked into an improvement of the excitation of the fluorophores. For this, a microscope equipped with an LED light source to improve the APC fluorophore excitation was used [24]. This resulted in a clear decrease of unidentified nucleated cells, see Figure 4.5. Also, to determine whether these remaining nuclei were enclosed by a cellular plasma membrane, we added wheat germ agglutinin to the CellSearch cocktail [5, 23]. This showed that part of the unidentified nucleated cells did not stain with wga, suggesting they are “bare”-like nuclei, see Figure 4.6. Whether or not these bare nuclei were actually present in the blood, or if they represent cells that have been damaged through the immunomagnetic selection or staining process, remains an outstanding question. However, a large fraction of leukocytes identified with the current immunostaining are not also stained with wga. This could suggest that the wga-staining assay is not sufficiently sensitive to stain all membranes or that the imaging method is not sufficiently sensitive to detect dim wga fluorescence. A high concentration of wga added to the CellSearch cocktail caused the aggregation of blood cells, and therefore a lower concentration was chosen for this assay. It might, however, be possible that a higher concentration of wga is necessary to locate all membranes present in the blood sample. Also, all samples were analyzed on the CellTracks, but improved identification will be observed when the LED light source is used. Another option is that the leukocyte membranes are damaged to a

point that the agglutinin can no longer bind to the residues on the membrane. This may be caused during the 15 min permeabilization process to open the membranes. The CellSearch permeabilization agent is saponin, a detergent which interacts with membrane cholesterols by selectively removing them to form pores, which enables antibodies to enter [19]. It is notable, that the fraction of the spiked MCF-7 cancer cells that were stained with wga was 100%. Compared to cancer cell line cells, the leukocytes are possibly more fragile and susceptible to damage caused during permeabilization. The wga+/DNA+/CD45–/CD16–/CK– cells are potentially CTCs that are not expressing CK for CellSearch detection. However, our observation that a similar population was detected in the blood of healthy volunteers makes this explanation less likely. Confirmation of whether or not CTCs are among the unidentified nucleated cells will likely have to come from DNA or RNA assays. Removing the cells from the cartridge might allow them to be selected based on their absence of fluorescence markers and subsequently sorted with fluorescence-activated cell sorting (FACS). Quantitative PCR after DNA isolation and amplification could reveal the presence of mutations, indicating one of the cells is, in fact, a tumor cell. Digital-droplet PCR has a higher sensitivity and could be applied on the entire cell suspension from a cartridge, although the multiplex analysis of mutations is less feasible in this assay. With these approaches, the number of tumor cells will have to be compared to the number of CTCs scored on the CellTracks (if present at all) to determine if more or an equal number of tumor cells are present in the cell suspension. Also, analysis of mRNA after FACS isolation might be feasible for the unidentified cells with a cellular membrane present. Quantification of a selection of proteins can indicate if the cells are of hematopoietic of epithelial lineage. However, the permeabilization of cells in the CellSearch might complicate this approach. Therefore, the CellSearch Epithelial Cell Profile Kit should be used. This kit is designed for RNA analysis and enriches EpCAM+ cells without subsequent permeabilization and staining. After taking several steps to determine the origin of unidentified cells, the possible identity for the cells that remained unidentified thus far might be:

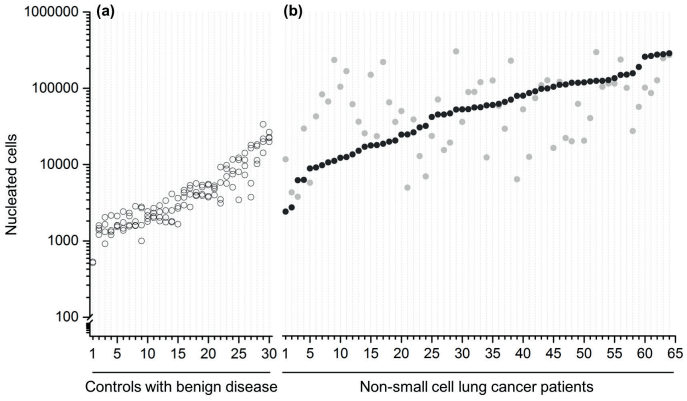
1. Hematopoietic cells not identified through damage incurred during the procedure or inaccessible antigens through the ferrofluids [9, 28];
2. Hematopoietic cells without sufficient expression of CD45 and CD16: This suggests they would need additional CD markers for identification or an improved labeling method that amplifies low signals, which might yield increased detection of very dim stained cells and separate the fluorescent signal of densely packed cells [14, 16, 33, 40]. Early myeloid cells have recently been observed to surround the tumor in high numbers, and as these cells do not yet express CD16, this remains a possibility [4];
3. CTCs with no or low expression of the CK antigens detected by the C11 and A53.B/A2 clones: Since these clones only recognize a subset of the CK present in a cell, it might be beneficial to include antibody clones that recognize all CK. Previously, we have shown that adding several CK clones to the CellSearch antibody cocktail improved the detection of CTC positive patients by 11% [12]. Also, it might be possible that EpCAM+/CK– CTCs are present, remaining undetected because of the downregulation of epithelial markers through epithelial-to-mesenchymal transition [21, 29, 34]. In or-

der to detect these cells, antibodies specific to this process could be added to the CellSearch immunostaining [3, 20, 35];

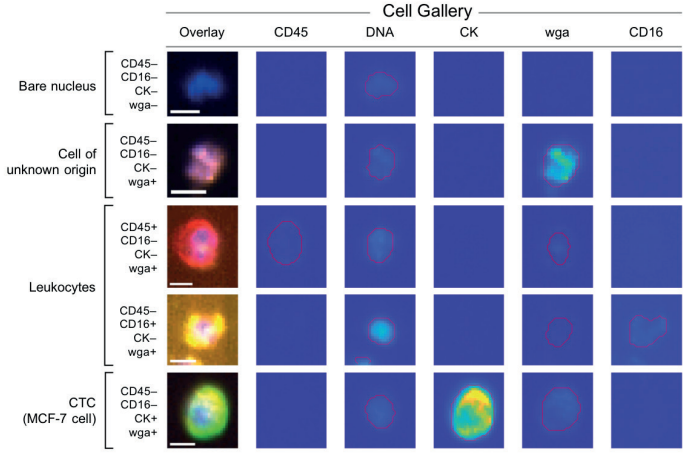
4. Cells of other origin: Such as circulating stromal, endothelial, or stem cells [6]. Detection of these cells would also require the addition of other antibodies to the assay.

Therefore, there is still much to be achieved toward the identification of the cells that remain anonymous. It is most important to determine if all tumor cells have been located or if some are still missed.

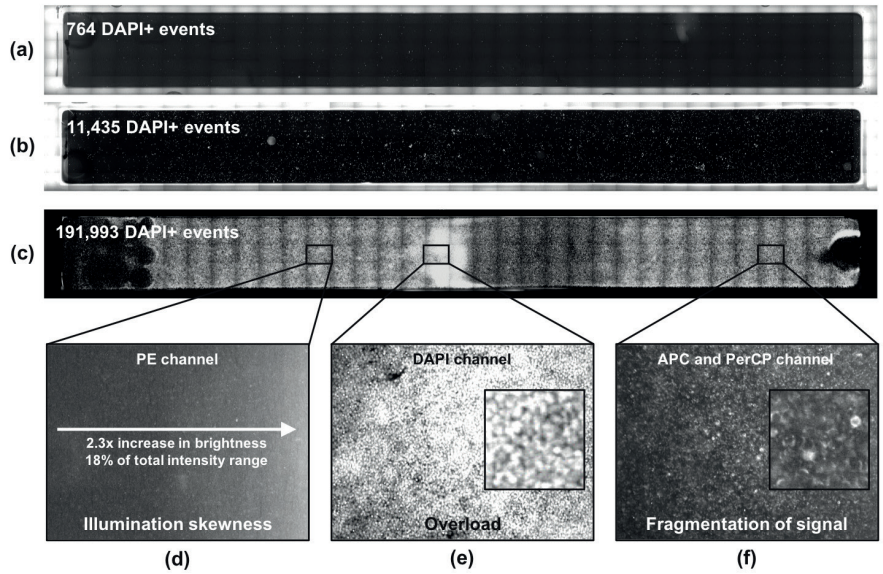
Conclusion Advanced image analysis was used to identify the cell populations present after immunomagnetic EpCAM enrichment by the CellSearch system from 659 cartridges from NSCLC patients. A significantly larger number of nucleated cells were found, from which the origin could not be identified, compared to samples from healthy volunteers and patients with benign disease. A large portion of the unidentified nucleated cells could be identified by: Adding CD16 to improve the identification of granulocytes; the use of an LED light source to improve fluorescence detection and; the addition of wga to discriminate between bare nuclei and cells. However, there is still much to gain in the identification of the cells that still remain anonymous. These observations must also be confirmed in patients with other cancers. The question remains: What are these cells? Going forward, it is most important to determine if all tumor cells have been located or if some are still missed and reside in this unidentified cell population. This unidentified CTC subpopulation could be of interesting clinical value and remains thus far unexploited. Therefore, further studies are needed to determine if EpCAM+/CK– CTCs are present among the nucleated cells of which the origin is not known.



Supplementary Figure 4.1: *Nucleated cell count in patients with benign disease and non-small cell lung cancer (NSCLC) patients.* Counts for four blood samples drawn at the same time in 30 patients with benign disease (a, open circle) and in 64 NSCLC patients with two blood samples (b): one drawn before initiation of therapy (right, dark circle) and one during therapy (light circle). For (a), the nucleated cell count in these samples was on average 5541 ± 541 and a median 3237; the coefficient of variation ranged from 9.9%–77.6% (mean 25.6% ± 14.6%). In (b), 46% of the 64 NSCLC patients showed an increase of more than 10% in nucleated cell count, in 31 (48%) a decrease of more than 10% and in 3 (5%) patients the change in the nucleated cell count was not more than 10%.



Supplementary Figure 4.2: *ACCEPT gallery of cells stained with the membrane lectin wheat germ agglutinin (wga).* The cell with wga-staining, but without the presence of any other markers, remains without determined origin, whereas the nucleus without wga-staining appears to be a bare nucleus. The scale bar is 10 pixels, representing 6.4 μm.



Supplementary Figure 4.3: *Illustration of the issues that influence the segmentation quality of the ACCEPT toolbox that were improved with a Deep Learning segmentation approach.* In total, 175 images of DAPI staining cover the surface of CellSearch cartridges for each sample. In (a) a cartridge containing less than 1000 DAPI+ events are shown; (b) shows a cartridge containing around 10,000 DAPI+ events and; (c) shows a cartridge containing much more than 100,000 DAPI+ events. While single cells can easily be detected in panels (a) and (b), the example in panel (c) demonstrates the effect that a high cell density makes the separation of cell boundaries nearly impossible. In (d) – (f) three of the 175 images of the very full cartridge are shown at a higher magnification to illustrate the three main difficulties for an image analysis program to identify objects. In (d) the illumination skewness of the PE channel corresponding to the DAPI image in (c) is presented. The background intensity, averaged over a small square of 5 by 5 pixels, is 2.3 times brighter on the right side of the image compared to the left one. In relation to the overall intensity range of 4095 intensity values, this is an increase of 18%. The segmentation model used assumes a nearly homogeneous background, but this assumption fails in case of an illumination bias. Panel (e) shows an overload of signal because the cartridge is too full with cells to detect cellular boundaries. Panel (f) shows the fragmentation of signal, which makes it for the software hard to accurately segment a cell and extract correct parameters, especially in combination with very packed areas as shown in panel (e). The segmentation results in case of overload (e) or fragmentation of signal (f) can be improved with a Deep Learning segmentation approach.

Bibliography

[1] W. L. Adair and S. Kornfeld. Isolation of the Receptors for Wheat Germ Agglutinin and the Ricinus Communis Lectins from Human Erythrocytes using Affinity Chromatography. *Journal of Biological Chemistry*, 249(15):4696–4704, 1974.

[2] W. J. Allard, J. Matera, M. C. Miller, M. I. Repollet, M. C. Connelly, C. G. Rao, A. G. Tibbe, J. W. Uhr, and L. W. Terstappen. Tumor Cells Circulate in the Peripheral Blood of All Major Carcinomas but not in Healthy Subjects or Patients with Nonmalignant Diseases. *Clinical Cancer Research*, 10(20):6897–6904, 2004.

[3] A. J. Armstrong, M. S. Marengo, S. Oltean, G. Kemeny, R. L. Bitting, J. D. Turnbull, C. I. Herold, P. K. Marcom, D. J. George, and M. A. Garcia-Blanco. Circulating Tumor Cells from Patients with Advanced Prostate and Breast Cancer Display Both Epithelial and Mesenchymal Markers. *Molecular Cancer Research*, 9(8):997–1007, 2011.

[4] A. Calcinotto, C. Spataro, E. Zagato, D. Di Mitri, V. Gil, M. Crespo, G. De Bernardis, M. Losa, M. Mirenda, E. Pasquini, A. Rinaldi, S. Sumanasuriya, M. B. Lambros, A. Neeb, R. Lucianò, C. A. Bravi, D. Nava-Rodrigues, D. Dolling, T. Prayer-Galetti, A. Ferreira, A. Briganti, A. Esposito, S. Barry, W. Yuan, A. Sharp, J. S. de Bono, and A. Alimonti. IL-23 Secreted by Myeloid Cells Drives Castration-Resistant Prostate Cancer. *Nature*, 559(7714):363, 2018.

[5] B. Chazotte. Labeling Membrane Glycoproteins or Glycolipids with Fluorescent Wheat Germ Agglutinin. *Cold Spring Harbor Protocols*, 2011(5):pdb–prot5623, 2011.

[6] Y. Chen, P. Li, P. H. Huang, Y. Xie, J. D. Mai, L. Wang, N. T. Nguyen, and T. J. Huang. Rare Cell Isolation and Analysis in Microfluidics. *Lab on a Chip*, 14(4):626–645, 2014.

[7] F. Chollet. Keras, 2015.

[8] S. J. Cohen, C. J. Punt, N. Iannotti, B. H. Saidman, K. D. Sabbath, N. Y. Gabrail, J. Picus, M. Morse, E. Mitchell, M. C. Miller, G. V. Doyle, H. Tissing, L. W. Terstappen, and N. J. Meropol. Relationship of Circulating Tumor Cells to Tumor Response, Progression-Free Survival, and Overall Survival in Patients with Metastatic Colorectal Cancer. *Journal of Clinical Oncology*, 26:3213–3221, 2008.

[9] F. A. Coumans and L. W. Terstappen. Detection and Characterization of Circulating Tumor Cells by the CellSearch Approach. *Methods in Molecular Biology*, 1347:263–278, 2015.

[10] M. Cristofanilli, G. T. Budd, M. J. Ellis, A. Stopeck, J. Matera, M. C. Miller, J. M. Reuben, G. V. Doyle, W. J. Allard, L. W. Terstappen, and D. F. Hayes. Circulating Tumor Cells, Disease Progression, and Survival in Metastatic Breast Cancer. *New England Journal of Medicine*, 351(8):781–791, 2004.

[11] J. S. de Bono, H. I. Scher, R. B. Montgomery, C. Parker, M. C. Miller, H. Tissing, G. V. Doyle, L. W. Terstappen, K. J. Pienta, and D. Raghavan. Circulating Tumor Cells Predict Survival Benefit From Treatment in Metastatic Castration-Resistant Prostate Cancer. *Clinical Cancer Research*, 14(19):6302–6309, 2008.

[12] S. de Wit, G. van Dalum, A. T. Lenferink, A. G. Tibbe, T. J. N. Hiltermann, H. J. Groen, C. J. van Rijn, and L. W. Terstappen. The Detection of EpCAM+ and EpCAM- Circulating Tumor Cells. *Scientific Reports*, 5:12270, 2015.

[13] B. Emde, A. Heinen, A. Gödecke, and K. Bottermann. Wheat Germ Agglutinin Staining as a Suitable Method for Detection and Quantification of Fibrosis in Cardiac Tissue after Myocardial Infarction. *European Journal of Histochemistry*, 58(4):2488, 2014.

[14] L. Faget and T. S. Hnasko. Tyramide Signal Amplification for Immunofluorescent Enhancement. *Methods in Molecular Biology*, 1318:161–172, 2015.

[15] P. Gazzaniga, A. Gradilone, E. de Berardinis, G. M. Busetto, C. Raimondi, O. Gandini, C. Nicolazzo, A. Petracca, B. Vincenzi, A. Farcomeni, V. Gentile, E. Cortesi, and L. Frati. Prognostic Value of Circulating Tumor Cells in Nonmuscle Invasive Bladder Cancer: a CellSearch Analysis. *Annals of Oncology*, 23(9):2352–2356, 2012.

[16] S. M. Gustafsdottir, E. Schallmeiner, S. Fredriksson, M. Gullberg, O. Söderberg, M. Jarvius, J. Jarvius, M. Howell, and U. Landegren. Proximity Ligation Assays for Sensitive and Specific Protein Analyses. *Analytical Biochemistry*, 345(1):2–9, 2005.

[17] T. J. N. Hiltermann, M. M. Pore, A. van den Berg, W. Timens, H. M. Boezen, J. J. Liesker, J. H. Schouwink, G. J. Wijnands, G. S. Kerner, F. A. Kruyt, H. Tissing, A. G. Tibbe, L. W. Terstappen, and H. J. Groen. Circulating Tumor Cells in Small-Cell Lung Cancer: a Predictive and Prognostic Factor. *Annals of Oncology*, 23(11):2937–2942, 2012.

[18] K. Hiraiwa, H. Takeuchi, H. Hasegawa, Y. Saikawa, K. Suda, T. Ando, K. Kumagai, T. Irino, T. Yoshikawa, S. Matsuda, M. Kitajima, and Y. Kitagawa. Clinical Significance of Circulating Tumor Cells in Blood from Patients with Gastrointestinal Cancers. *Annals of Surgical Oncology*, 15(11):3092, 2008.

[19] M. C. Jamur and C. Oliver. Immunocytochemical Methods and Protocols - Permeabilization of Cell Membranes. *Methods in Molecular Biology*, 588:63–66, 2010.

[20] G. Kallergi, M. A. Papadaki, E. Politaki, D. Mavroudis, V. Georgoulias, and S. Agelaki. Epithelial to Mesenchymal Transition Markers Expressed in Circulating Tumour Cells of Early and Metastatic Breast Cancer Patients. *Breast Cancer Research*, 13(3):R59, 2011.

[21] M. G. Krebs, J.-M. Hou, R. Sloane, L. Lancashire, L. Priest, D. Nonaka, T. H. Ward, A. Backen, G. Clack, A. Hughes, M. Ranson, F. H. Blackhall, and C. Dive. Analysis of Circulating Tumor Cells in Patients with Non-Small Cell Lung Cancer Using Epithelial Marker-Dependent and -Independent Approaches. *Journal of Thoracic Oncology*, 7(2):306–315, 2012.

[22] M. G. Krebs, R. Sloane, L. Priest, L. Lancashire, J.-M. Hou, A. Greystoke, T. H. Ward, R. Ferraldeschi, A. Hughes, G. Clack, M. Ranson, C. Dive, and F. H. Blackhall. Evaluation and Prognostic Significance of Circulating Tumor Cells in Patients with Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology*, 29(12):1556–1563, 2011.

[23] W. R. Legant, L. Shao, J. B. Grimm, T. A. Brown, D. E. Milkie, B. B. Avants, L. D. Lavis, and E. Betzig. High-Density Three-Dimensional Localization Microscopy Across Large Volumes. *Nature Methods*, 13(4):359, 2016.

[24] Lumencor. I. Solar Light Engine.

[25] E. Racila, D. Euhus, A. J. Weiss, C. G. Rao, J. McConnell, L. W. Terstappen, and J. W. Uhr. Detection and Characterization of Carcinoma Cells in the Blood. *Proceedings of the National Academy of Sciences of the United States of America*, 95(8):4589–4594, 1998.

[26] C. G. Rao, T. Bui, M. C. Connelly, G. V. Doyle, I. Karydis, M. R. Middleton, G. Clack, M. Malone, F. A. Coumans, and L. W. Terstappen. Circulating Melanoma Cells and Survival in Metastatic Melanoma. *International Journal of Oncology*, 38(3):755–760, 2011.

[27] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[28] T. M. Scholtens. *Automated Classification and Enhanced Characterization of Circulating Tumor Cells by Image Cytometry*. PhD thesis, University of Twente, 2012.

[29] M. J. Serrano, F. G. Ortega, M. J. Alvarez-Cubero, R. Nadal, P. Sanchez-Rovira, M. Salido, M. Rodríguez, J. L. García-Puche, M. Delgado-Rodriguez, F. Solé, M. A. García, M. Perán, R. Rosell, J. A. Marchal, and J. A. Lorente. EMT and EGFR in CTCs Cytokeratin Negative Non-Metastatic Breast Cancer. *Oncotarget*, 5(17):7486, 2014.

[30] N. Sharon and H. Lis. *Lectins: Second Edition*. 2007.

[31] C. Solórzano, S. Bouquelet, M. A. Pereyra, F. Blanco-Favela, M. C. Slomianny, R. Chavez, R. Lascurain, E. Zenteno, and C. Agundis. Isolation and Characterization of the Potential Receptor for Wheat Germ Agglutinin from Human Neutrophils. *Glycoconjugate Journal*, 23(7-8):591–598, 2006.

[32] G. van Dalum, G.-J. Stam, L. F. Scholten, W. J. Mastboom, I. Vermes, A. G. Tibbe, M. R. de Groot, and L. W. Terstappen. Importance of Circulating Tumor Cells in Newly Diagnosed Colorectal Cancer. *International Journal of Oncology*, 46(3):1361–1368, 2015.

[33] G. van Dalum, G. J. van der Stam, A. G. Tibbe, B. Franken, W. J. Mastboom, I. Vermes, M. R. de Groot, and L. W. Terstappen. Circulating Tumor Cells Before and During Follow-Up After Breast Cancer Surgery. *International Journal of Oncology*, 46(1):407–413, 2015.

[34] B. Willipinski-Stapelfeldt, S. Riethdorf, V. Assmann, U. Woelfle, T. Rau, G. Sauter, J. Heukeshoven, and K. Pantel. Changes in Cytoskeletal Protein Composition Indicative of an Epithelial-Mesenchymal Transition in Human Micrometastatic and Primary Breast Carcinoma Cells. *Clinical Cancer Research*, 11(22):8006–8014, 2005.

[35] S. Wu, S. Liu, Z. Liu, J. Huang, X. Pu, J. Li, D. Yang, H. Deng, N. Yang, and J. Xu. Classification of Circulating Tumor Cells by Epithelial-Mesenchymal Transition Markers. *PLOS ONE*, 10(4):e0123976, 2015.

[36] L. L. Zeune. Deep Learning to Identify Circulating Tumor Cells by ACCEPT, 2018.

[37] L. L. Zeune, S. de Wit, A. S. Berghuis, M. J. IJzerman, L. W. Terstappen, and C. Brune. How to Agree on a CTC. *Cytometry Part A*, 2018.

[38] L. L. Zeune, G. van Dalum, C. Decraene, C. Proudhon, T. Fehm, H. Neubauer, B. Rack, M. Alunni-Fabbroni, L. W. Terstappen, S. A. van Gils, and C. Brune. Quantifying HER-2 Expression on Circulating Tumor Cells by ACCEPT. *PLOS ONE*, 12(10):e0186562, oct 2017.

[39] L. L. Zeune, G. van Dalum, L. W. Terstappen, S. A. van Gils, and C. Brune. Multiscale Segmentation via Bregman Distances and Nonlinear Spectral Analysis. *SIAM Journal on Imaging Sciences*, 10(1):111–146, 2017.

[40] S. Zhou, L. Yuan, X. Hua, L. Xu, and S. Liu. Signal Amplification Strategies for DNA and Protein Detection Based on Polymeric Nanocomposites and Polymerization: A Review. *Analytica Chimica Acta*, 877:19–32, 2015.

Chapter 5

How to Agree on a CTC

Evaluating the Consensus in Circulating Tumor Cell Scoring

Leonie L. Zeune*, Sanne de Wit*, A.M. Sofie Berghuis,
Maarten J. IJzerman, Leon W.M.M. Terstappen and Christoph Brune

Cytometry Part A 93.12 (2018): 1202-1206

* Both authors contributed equally.

Abstract

For using counts of circulating tumor cells (CTC) in the clinic to aid a physician's decision, its reported values will need to be accurate and comparable between institutions. Many technologies have become available to enumerate and characterize CTCs, thereby showing a large range of reported values. Here we introduce an Open Source CTC scoring tool to enable comparison of different reviewers and facilitate the reach of a consensus on assigning objects as CTCs. One hundred images generated from two different platforms were used to assess concordance between 15 reviewers and an expert panel. Large differences were observed between reviewers in assigning objects as CTCs urging the need for computer recognition of CTCs. A demonstration of a Deep Learning approach on the 100 images showed the promise of this technique for future CTC enumeration.

5.1 Introduction

The peripheral blood load of Circulating Tumor Cells (CTC) enumerated with the CellSearch® system (Menarini Silicon Biosystems Inc, Huntingdon Valley, PA, USA) system is directly related to the survival prospects of patients with metastatic cancer and their presence in pa-

tients with primary cancers is related to an increased risk at disease recurrence as well as survival [1, 4, 7, 8, 11, 13, 15, 23]. To monitor therapy based on CTC counts and to diagnose the presence of cancer beyond the primary tumor, it is of utmost importance to accurately assign objects as CTCs. To determine whether a CTC count is below or above 5 CTCs the inter-reader variability is challenging, but can be quite low [12, 14, 22]. However, when a significant change in CTC number will need to be assessed, the challenge increases [6]. Yet, nearly all CTC isolation techniques lack a fully automated image analysis, thereby making CTC counts subjective to the reviewer. Here we evaluate the consensus between multiple reviewers in assigning objects as CTCs by introducing an Open Source Scoring Tool that enables comparisons between reviewers and improve their ability to reach consensus (https://github.com/leoniez/ACCEPT/releases). Finally, we introduce a fully automated CTC classification system based on Deep Learning (DL).

5.2 Methods

5.2.1 Selection of Fluorescent Cell Images for CTC Scoring

A set of 100 fluorescent cell images comprising of potential CTCs and leukocytes was assembled. The fluorescent images were obtained from 7.5 mL blood samples from metastatic prostate and non-small cell lung cancer patients [9, 10, 21]. The blood was processed with the CellSearch® system and the EpCAM depleted blood was collected, passed through microsieves, fluorescently labeled and images taken by fluorescence microscopy [10]. The images exported from the CellSearch® system as well as fluorescent images after filtration of the EpCAM depleted blood were all reanalyzed with the ACCEPT toolbox (which can be downloaded from https://github.com/LeonieZ/ACCEPT). 100 thumbnail images were selected from which 50 were obtained from CellSearch® images and 50 from the microsieves.

5.2.2 Development of the ACCEPT Scoring Tool

For the cell scoring we developed the ACCEPT CTC Scoring tool in MATLAB® 2016a (Mathworks, Natick MA, USA). This tool will be released as part of the general ACCEPT toolbox for CTC analysis in a future release. A screenshot of the tool is shown in Figure 5.1. For every object, the thumbnail images of three fluorescent channels (CD45, DAPI, PE) are shown using the full range of color intensities. The red contour surrounding the object indicates its boundary, which is automatically detected by an advanced segmentation algorithm in ACCEPT [25, 26]. Next to it, an overlay image is presented, showing the CD45 signal in red, the DAPI signal in blue and the PE signal in green. Here, the intensity is scaled from the smallest to the largest intensity value found inside the red contour. The right three scatter plots show six quantitative measurements exported from the objects to facilitate the scoring. A reviewer can choose from four presented answers: “1. Definitely not a CTC”; “2. Most likely not a CTC”; “3. Most likely a CTC”; and “4. Definitely a CTC”. Once the user decides for one answer, the tool will automatically proceed to the next object and it is not possible to change an answer

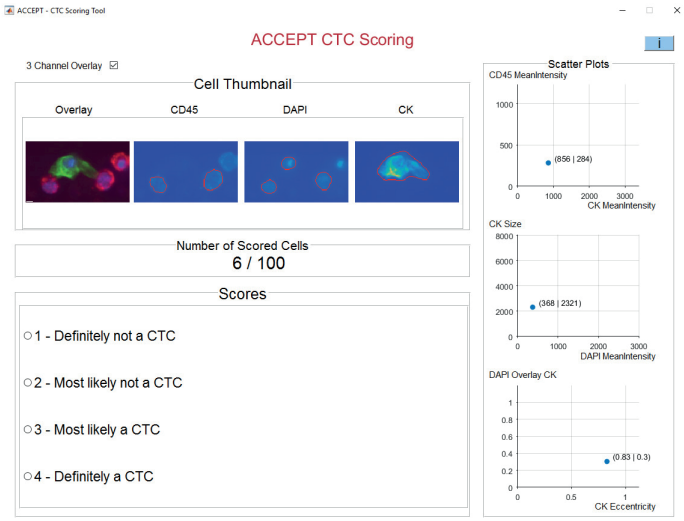


Figure 5.1: *Screenshot of the ACCEPT CTC Scoring Tool.* For each presented cell a thumbnail gallery of all fluorescent channels is shown with four answers and three plots presenting measurement information of the respective cell to aid in the decision. After selecting an answer, the program automatically proceeds to the next cell.

afterwards. All objects are presented in a randomized order. With this tool, 15 independent reviewers from 6 different institutes scored the set of 100 cells. These scores formed the basis of our analysis.

5.2.3 Developing the "Ground-Truth"

One of the major drawbacks of image-based CTC analysis is the lack of a ground-truth solution. While molecular information could be used to determine if a cell really is a CTC, this information is in most of the cases not available or very hard to acquire. To compensate for the lack of a real ground-truth answer for our set, we formed an Expert Panel (EP) consisting of four experts in the field and had them score the 100 objects together in one session. Two of the experts were involved in the original definition of a CTC in the CellSearch® system and the other two were trained reviewers with several years of experience in scoring CTCs. All 100 objects were discussed in an online meeting and they had to agree on one answer for each object. These answers were used as a ground-truth for comparison with the result of the average reviewer (AR) and DL.

5.2.4 Automated CTC Classification by Deep Learning

A DL network approach was used to classify the 100 objects. The classifier was trained on images obtained from the CellSearch® system and contained 13,123 CTC candidates, 18,820

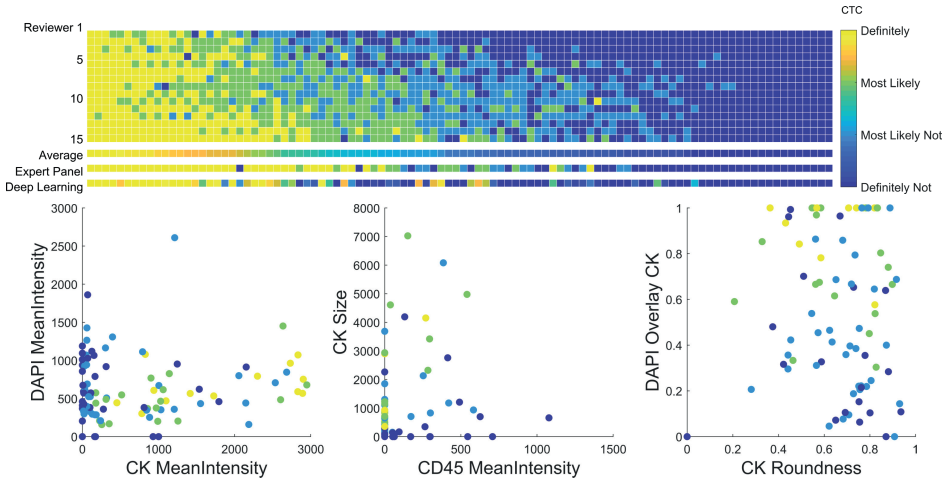


Figure 5.2: *Summary of the CTC Scoring Results.* Results by 15 reviewers, summarized with the Average Reviewer, followed by the results of the Expert Panel and the Deep Learning automated CTC scoring (upper panel). The Average Reviewer score of all 100 cells are presented in three scatter plots using several parameters: the mean intensity of the signal detected in all three channels (DAPI, CK and CD45), the size and roundness of the CK signal and the overlay between the DAPI and CK signals (lower panel).

objects with DNA staining not classified as CTCs and 8,548 other objects. The details of the DL approach are described elsewhere [24].

5.2.5 Statistical Analysis

To rule out any agreement by chance, inter-rater agreement between all 15 reviewers was determined with Fleiss' Kappa κ, whereas intra-rater agreements between the AR, EP and DL were determined with Cohen's kappa κ. The agreement was considered poor if $\kappa \leq 0.20$, fair for κ 0.21-0.40, moderate for κ 0.41-0.60, substantial or good for κ 0.61-0.80 and almost perfect for κ 0.81-1.00 [2, 16].

5.3 Results

5.3.1 Scores

The scores of all 100 objects from each reviewer, the AR score, the EP score and the DL score are presented in Figure 5.2. The probability for a cell being a "CTC" returned from the DL network is scaled from 0.0 (displayed as score 1 by reviewers, visualized as dark blue) to 1.0 (displayed as score 4, visualized as yellow). Images of all 100 objects including the AR, EP and DL score are provided in Supplementary Figure 5.1.

5.3.2 Consensus between Reviewers

Only one object out of 100 objects was scored as "Definitely a CTC" by all 15 investigators, and for only 13 more objects everyone agreed on the answer (all "Definitely not a CTC"). If we summarize answers in two classes: a "CTC" class (objects scored as 3 or 4) and "Not a CTC" class (objects scored as 1 or 2), these numbers increased to 11 objects scored as "CTC" and 30 objects scored as "Not a CTC" by all users, yet thereby 59 objects remain ambiguous. Of these 41 objects where everyone agreed on the same class, 54% were cells extracted from CellSearch® and 46% were cells extracted from microsieve filtration. In total 1500 scores were assigned; 42% was for "Definitely not a CTC", 22% for "Most likely not a CTC", 19% for "Most likely a CTC" and 17% for "Definitely a CTC". The inter-rater agreement was calculated with Fleiss' kappa and this shows a fair agreement with κ 0.38 for the case of four possible answers yet a moderate agreement with κ 0.60 if we summarize the answers in two classes.

5.3.3 Consensus of Expert Panel with Average Reviewer

When the results were divided into "CTC" and "Not a CTC", the overall agreement between EP and AR score was 80% (Table 5.1 (a)). The agreement between single reviewers and the EP ranged from 70% to 88%. When the EP scored "Not a CTC", the AR agreed in 98% of the cases. When a "CTC" was scored according to the EP, the reviewers agreed only in 61%, showing that the consensus for a cell being "Not a CTC" is much higher than a positive agreement. Cohen's kappa was calculated to determine the intra-agreement, which showed a moderate agreement of κ 0.60. The agreement in case of four possible answers is summarized in Supplemental Table 5.1 (a).

5.3.4 Consensus with Deep Learning

The same set of 100 objects was reviewed by a DL network. Compared with AR, there was an overall agreement of 84%, using the two classes (see Table 5.1 (b)). For objects that were scored as "Not a CTC" by the DL, the AR agreed in 92% of the cases, whereas the AR agreed on objects that were scored as "CTC" in 70%. This is in contrast to the comparison of the EP scores with DL (see Table 5.1 (c)). Here, the agreement with DL is 71% of the objects scored as "Not a CTC", but 84% on objects scored as "CTC". The agreement between AR, EP and DL for all four classes can be found in Supplementary Table 5.1. Notably, the AR scored 13 cells as "Definitely a CTC" (score 4) and in all 13 cases both the EP and DL also scored as "Definitely a CTC". Intra-agreement with Cohen's kappa for AR with DL was κ 0.64 and for EP with DL was κ 0.52, which can be interpreted as a substantial (AR to DL) and moderate (EP to DL) agreement. Calculating Fleiss' kappa for agreement between AR, EP and DL, showed a κ of 0.58 for two classes, which can be considered a moderate agreement as well.

Agreement: 80 $\kappa = 0.60$		**Expert Panel** Not a CTC	CTC
Average Reviewer	Not a CTC	50	19
	CTC	1	30

(a)

Agreement: 84 $\kappa = 0.64$		**Deep Learning** Not a CTC	CTC
Average Reviewer	Not a CTC	58	11
	CTC	5	26

(b)

Agreement: 76 $\kappa = 0.52$		**Deep Learning** Not a CTC	CTC
Expert Panel	Not a CTC	45	6
	CTC	18	31

(c)

Table 5.1: Overview of agreement on 100 cells between (a) Average Reviewer and Expert Panel, (b) Average Reviewer and Deep Learning and (c) Expert Panel and Deep Learning, summarizing scores as a “CTC” class and “Not a CTC” class.

5.4 Discussion

The demonstration that the peripheral blood tumor cell load is directly related to the clinical outcome of the patients, has led to the introduction of various platforms to enumerate and characterize CTCs [3] (refs will be added from the technologies in this issue). The morphological appearance of these CTCs is however extremely heterogeneous [5, 17, 18]. This makes it difficult to arrive at a common definition of what is and what is not a "CTC". Although genetic aberrations in the detected cells can confirm that the object indeed is a cancer cell, in practice this cannot be performed on all the CTC candidates. Moreover, the largest fraction of CTCs or CTC related objects are either undergoing apoptosis or are extracellular vesicles derived from the cancer cells and might not have the complete genome represented [17, 20]. However, the presence of these tumor related events is also related to poor clinical outcome [5, 19]. To report a CTC count for the disease management of cancer patients, one will need to be able to rely on the truthfulness of the count. This is not as straightforward as one might think, since all the technologies presenting CTC data report different numbers of CTC and little data is available on the variability of assigning objects as CTC. Here, we introduce a tool which can help assessing the concordance between reviewers on assigning object as CTC. It can also be used to train the reviewers to increase concordance in reaching consensus on which objects are assigned as CTC and which ones are not. In this study, we evaluate the consensus between 15 reviewers and an Expert Panel in assigning 100 objects as CTC, using an ACCEPT Open Source Scoring Tool (https://github.com/leoniez/ACCEPT/releases). Agreement between reviewers on scoring all objects in four categories resulted in quite a large variation, as shown in Figure 5.2. No significant difference was observed between the images originating from CellSearch® or from microsieve filtration. Previously, we have shown that the concordance between reviewers to score the expression of Her-2 on CTC can clearly be improved by the use of the ACCEPT toolbox [25]. Although using a "yes" or "no" for CTC scoring improves the agreement (see Table 5.1 and Supplementary Table 5.1), the large heterogeneity of CTC morphology makes it virtually impossible to obtain a perfect agreement between reviewers. In comparison to previous studies evaluating the consensus in scoring cells as CTC we reported lower values for Fleiss' Kappa κ [12, 14]. The reasons for this deviation are versatile. First, the visualization used in the ACCEPT toolbox is different to the visualization in the CellSearch® system and reviewers are less used to the new visualization. ACCEPT shows the true value of staining in the images, in contrast to a scaled-up representation of the staining in CellSearch®. Moreover, if we allow four possible answers the probability of agreement decreases and lowers the Fleiss' Kappa. The composition of the training set is also an important factor. We have seen that the agreement of a cell being not a CTC tends to be higher than the agreement on being a CTC. Thus, a set consisting of a lot of negative examples is more likely to result in a high agreement [14]. In our set, we tried to balance the number of positive and negative examples to account for that influence. In another comparison study cells that resulted in high disagreement in their expert panel were excluded from further analysis which most probably would have let to a lower agreement [12]. Contrary to their approach, we aimed to include objects that were not directly obvious to classify as a CTC or not, since these type of cells will be most frequently present in the

clinic. Automated image analysis can eliminate any variation, since a computer will always use the same rationale to reach a conclusion. Therefore, we used a Deep Learning network that was trained on 40,491 thumbnail images obtained from CellSearch® to identify CTC. For each object, the Deep Learning network provides a likelihood ratio whether the object is a CTC or not. The network performed remarkably well on the images of the 100 objects and no real difference could be observed between images obtained from CellSearch® or from microsieve filtration (see Table 5.1 and Figure 5.2). The images available from other platforms are however limited, and the future will tell whether this Deep Learning network can identify CTC independent from the platform they have been generated from or if the network needs to be trained on images originating from various platforms. In conclusion, we have shown that CTC agreement between reviewers can vary greatly and this presents a complication for using true CTC counts in the clinic. To reach objective agreement on CTC scoring, automated image analysis might hold the answer. We invite you to use the toolbox presented here to determine the rater variability in CTC scoring at your laboratory.

Acknowledgements

We are grateful for the expert panel consisting of Madeline Repollet, Joost Swennenhuis, Frances Tanney and Leon Terstappen and the reviewers Marianna Alunni-Fabbroni, Kiki Andree, Mateus Crespo, Agustin Enciso Martinez, Penny Flohr, Rita Lampignano, Leonie Majunke, Mariangela Manicone, Anouk Mentink-Leusink, Afroditi Nanou, Marianne Oulhen, Elisabetta Rossi, Liwen Yang and Beate Zill for scoring the images using the ACCEPT tool.

Agreement: 53 $\kappa = 0.38$		**Expert Panel** 1	2	3	4	**Total**
Average Reviewer	1	30	3	0	1	**34**
	2	8	9	11	7	**35**
	3	1	0	1	16	**18**
	4	0	0	0	13	**13**
	Total	**39**	**12**	**12**	**37**	**100**

(a)

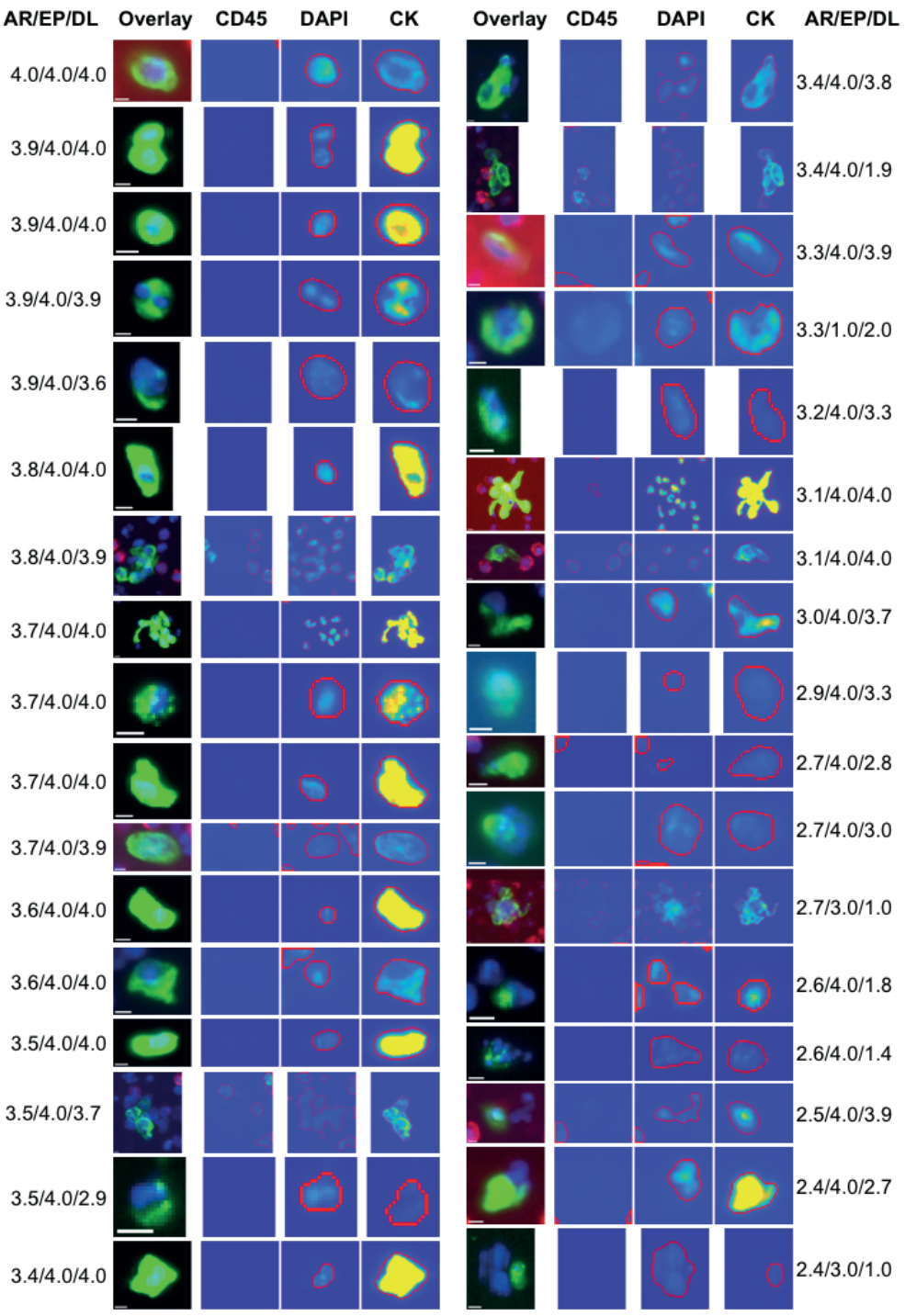
Agreement:52 $\kappa = 0.34$		**Deep Learning** 1	2	3	4	**Total**
Average Reviewer	1	30	3	1	0	**34**
	2	21	4	8	2	**35**
	3	2	3	5	8	**18**
	4	0	0	0	13	**13**
	Total	**53**	**10**	**14**	**23**	**100**

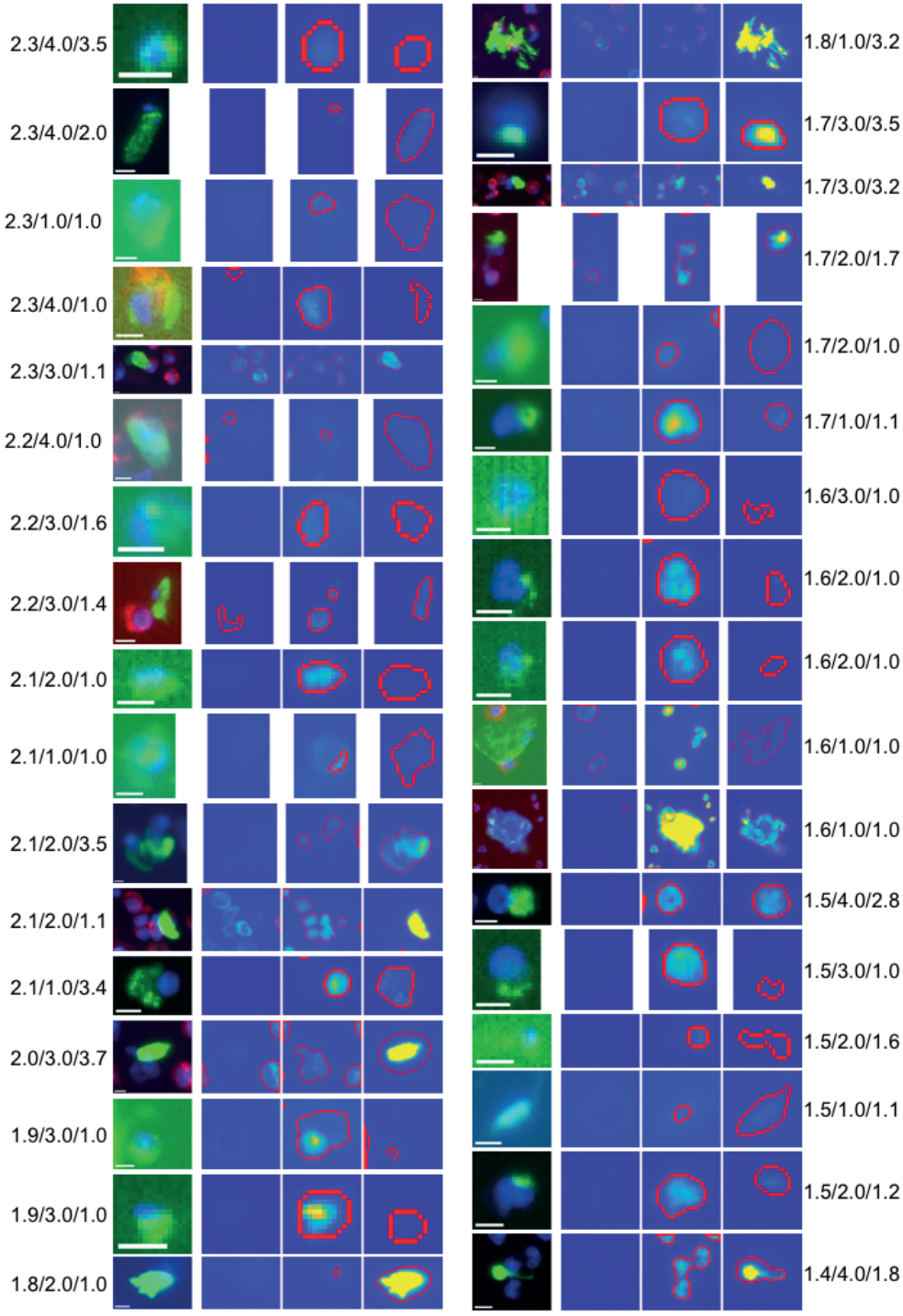
(b)

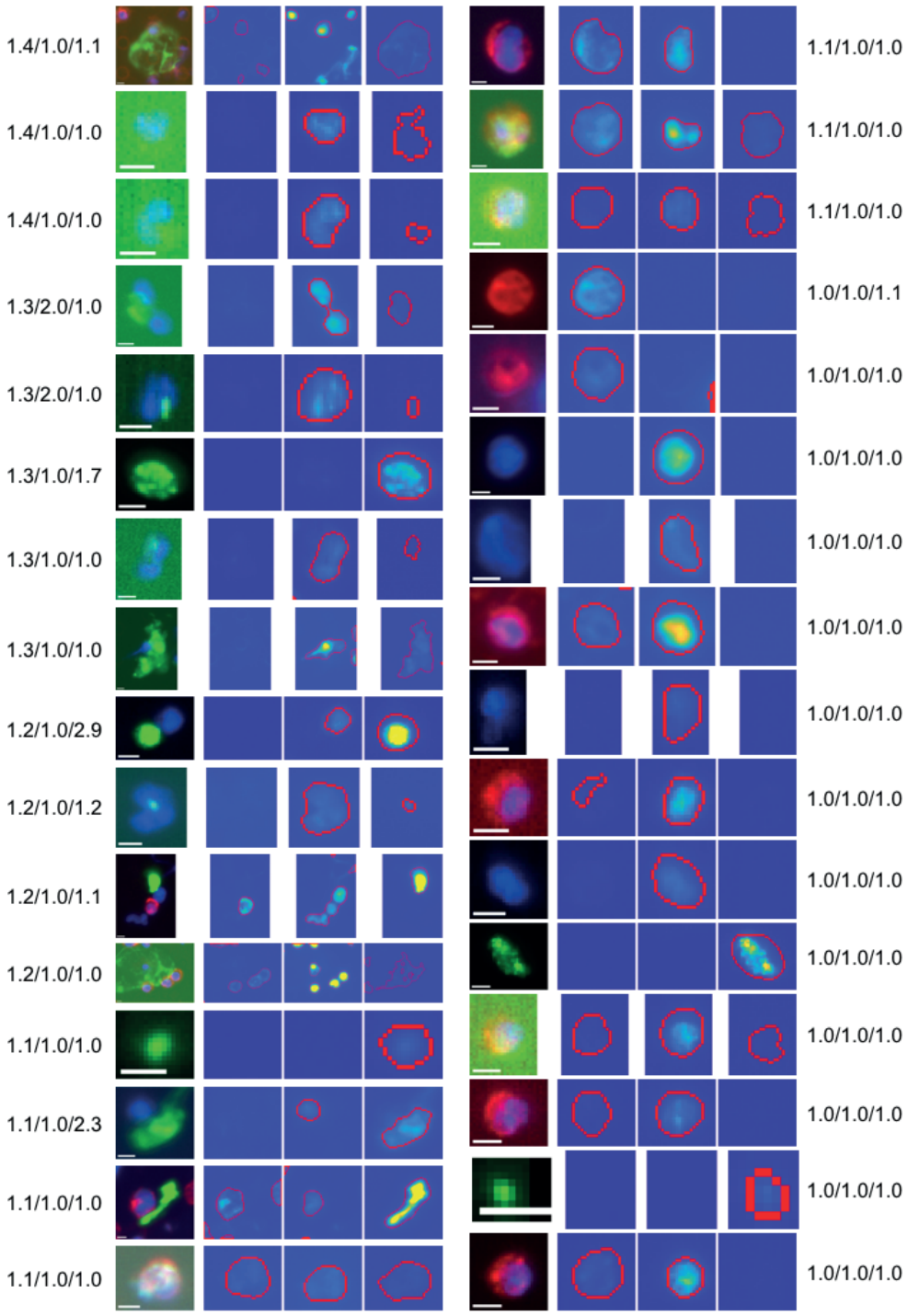
Agreement: 59 $\kappa = 0.40$		**Deep Learning** 1	2	3	4	**Total**
Expert Panel	1	34	3	2	0	**39**
	2	7	1	4	0	**12**
	3	7	2	2	1	**12**
	4	5	4	6	22	**37**
	Total	**53**	**10**	**14**	**23**	**100**

(c)

Supplementary Table 5.1: Overview of agreement on 100 cells between (a) Average Reviewer and Expert Panel; (b) Average Reviewer and Deep Learning and; (c) Expert Panel and Deep Learning scored with "1. Definitely not a CTC"; "2. Most likely not a CTC"; "3. Most likely a CTC"; and "4. Definitely a CTC".


AR/EP/DL
Overlay
CD45
DAPI
CK
Overlay
CD45
DAPI
CK
AR/EP/DL
4.0/4.0/4.0
3.9/4.0/4.0
3.9/4.0/4.0
3.9/4.0/3.9
3.9/4.0/3.6
3.8/4.0/4.0
3.8/4.0/3.9
3.7/4.0/4.0
3.7/4.0/4.0
3.7/4.0/4.0
3.7/4.0/3.9
3.6/4.0/4.0
3.6/4.0/4.0
3.5/4.0/4.0
3.5/4.0/3.7
3.5/4.0/2.9
3.4/4.0/4.0
3.4/4.0/3.8
3.4/4.0/1.9
3.3/4.0/3.9
3.3/1.0/2.0
3.2/4.0/3.3
3.1/4.0/4.0
3.1/4.0/4.0
3.0/4.0/3.7
2.9/4.0/3.3
2.7/4.0/2.8
2.7/4.0/3.0
2.7/3.0/1.0
2.6/4.0/1.8
2.6/4.0/1.4
2.5/4.0/3.9
2.4/4.0/2.7
2.4/3.0/1.0


2.3/4.0/3.5
2.3/4.0/2.0
2.3/1.0/1.0
2.3/4.0/1.0
2.3/3.0/1.1
2.2/4.0/1.0
2.2/3.0/1.6
2.2/3.0/1.4
2.1/2.0/1.0
2.1/1.0/1.0
2.1/2.0/3.5
2.1/2.0/1.1
2.1/1.0/3.4
2.0/3.0/3.7
1.9/3.0/1.0
1.9/3.0/1.0
1.8/2.0/1.0
1.8/1.0/3.2
1.7/3.0/3.5
1.7/3.0/3.2
1.7/2.0/1.7
1.7/2.0/1.0
1.7/1.0/1.1
1.6/3.0/1.0
1.6/2.0/1.0
1.6/2.0/1.0
1.6/1.0/1.0
1.6/1.0/1.0
1.5/4.0/2.8
1.5/3.0/1.0
1.5/2.0/1.6
1.5/1.0/1.1
1.5/2.0/1.2
1.4/4.0/1.8



Supplementary Figure 5.1: *Gallery of the 100 objects that were presented to the 15 reviewers, Expert Panel and Deep Learning network for scoring.* The objects are sorted according to the AR score (high to low), followed by EP (high to low), followed by DL (high to low). The AR, EP and DL score is indicated next to the images.

Bibliography

[1] W. J. Allard, J. Matera, M. C. Miller, M. I. Repollet, M. C. Connelly, C. G. Rao, A. G. Tibbe, J. W. Uhr, and L. W. Terstappen. Tumor Cells Circulate in the Peripheral Blood of All Major Carcinomas but not in Healthy Subjects or Patients with Nonmalignant Diseases. *Clinical Cancer Research*, 10(20):6897–6904, 2004.

[2] D. G. Altman. *Practical Statistics for Medical Research*. CRC press, 1991.

[3] A. M. Barradas and L. W. Terstappen. Towards the Biological Understanding of CTC: Capture Technologies, Definitions and Potential to Create Metastasis. *Cancers*, 5(4):1619–1642, 2013.

[4] S. J. Cohen, C. J. Punt, N. Iannotti, B. H. Saidman, K. D. Sabbath, N. Y. Gabrail, J. Picus, M. Morse, E. Mitchell, M. C. Miller, G. V. Doyle, H. Tissing, L. W. Terstappen, and N. J. Meropol. Relationship of Circulating Tumor Cells to Tumor Response, Progression-Free Survival, and Overall Survival in Patients with Metastatic Colorectal Cancer. *Journal of Clinical Oncology*, 26:3213–3221, 2008.

[5] F. A. Coumans, C. J. Doggen, G. Attard, J. S. de Bono, and L. W. Terstappen. All Circulating EpCAM+ CK+ CD45- Objects Predict Overall Survival in Castration-Resistant Prostate Cancer. *Annals of Oncology*, 21(9):1851 – 1857, 2010.

[6] F. A. Coumans, S. T. Ligthart, and L. W. Terstappen. Interpretation of Changes in Circulating Tumor Cell Counts. *Translational Oncology*, 5(6):486–491, 2012.

[7] M. Cristofanilli, G. T. Budd, M. J. Ellis, A. Stopeck, J. Matera, M. C. Miller, J. M. Reuben, G. V. Doyle, W. J. Allard, L. W. Terstappen, and D. F. Hayes. Circulating Tumor Cells, Disease Progression, and Survival in Metastatic Breast Cancer. *New England Journal of Medicine*, 351(8):781–791, 2004.

[8] J. S. de Bono, H. I. Scher, R. B. Montgomery, C. Parker, M. C. Miller, H. Tissing, G. V. Doyle, L. W. Terstappen, K. J. Pienta, and D. Raghavan. Circulating Tumor Cells Predict Survival Benefit From Treatment in Metastatic Castration-Resistant Prostate Cancer. *Clinical Cancer Research*, 14(19):6302–6309, 2008.

[9] S. de Wit. *Circulating Tumor Cells and Beyond*. PhD thesis, University of Twente, Enschede, The Netherlands, jun 2018.

[10] S. de Wit, M. Manicone, E. Rossi, R. Lampignano, L. Yang, B. Zill, A. Rengel-Puertas, M. Oulhen, M. Crespo, A. S. Berghuis, K. C. Andree, R. Vidotto, E. K. Trapp, M. Tzschaschel, E. Colomba, G. Fowler, P. Flohr, P. Rescigno, M. Sousa Fontes, R. Zamarchi, T. Fehm, H. Neubauer, B. Rack, M. Alunni-Fabbroni, F. Farace, J. S. de Bono, M. J. IJzerman, and L. W. Terstappen. EpCAMhigh and EpCAMlow Circulating Tumour Cells in Metastatic Prostate and Breast Cancer Patients. 2017.

[11] T. J. N. Hiltermann, M. M. Pore, A. van den Berg, W. Timens, H. M. Boezen, J. J. Liesker, J. H. Schouwink, G. J. Wijnands, G. S. Kerner, F. A. Kruyt, H. Tissing, A. G. Tibbe, L. W. Terstappen, and H. J. Groen. Circulating Tumor Cells in Small-Cell Lung Cancer: a Predictive and Prognostic Factor. *Annals of Oncology*, 23(11):2937–2942, 2012.

[12] M. Ignatiadis, S. Riethdorf, F.-C. Bidard, I. Vaucher, M. Khazour, F. Rothé, J. Metallo, G. Rouas, R. E. Payne, R. Coombes, I. Teufel, U. Andergassen, S. Apostolaki, E. Politaki, D. Mavroudis, S. Bessi, M. Pestrin, A. Di Leo, M. Campion, M. Reinholz, E. Perez, M. Piccart, E. Borgen, B. Naume, J. Jimenez, C. Aura, L. Zorzino, M. Cassatella, M. Sandri, B. Mostert, S. Sleijfer, J. Kraan, W. J. Janni, T. Fehm, B. Rack, L. W. Terstappen, M. I. Repollet, J.-Y. Pierga, M. C. Miller, C. Sotiriou, S. Michiels, and K. Pantel. International Study on Inter-Reader Variability for Circulating Tumor Cells in Breast Cancer. *Breast Cancer Research*, 16(2):R43, 2014.

[13] W. J. Janni, B. Rack, L. W. Terstappen, J.-Y. Pierga, F.-A. Taran, T. Fehm, C. Hall, M. R. de Groot, F.-C. Bidard, T. W. Friedl, P. A. Fasching, S. Y. Brucker, K. Pantel, and A. Lucci. Pooled Analysis of the Prognostic Relevance of Circulating Tumor Cells in Primary Breast Cancer. *Clinical Cancer Research*, 22(10):2583–2593, 2016.

[14] J. Kraan, S. Sleijfer, M. H. Strijbos, M. Ignatiadis, D. Peeters, J.-Y. Pierga, F. Farace, S. Riethdorf, T. Fehm, L. Zorzino, A. G. Tibbe, M. Maestro, R. Gisbert-Criado, G. Denton, J. S. de Bono, C. Dive, J. A. Foekens, and J. W. Gratama. External Quality Assurance of Circulating Tumor Cell Enumeration Using the CellSearch® System: A Feasibility Study. *Cytometry Part B*, 80B(2):112–118, 2011.

[15] M. G. Krebs, R. Sloane, L. Priest, L. Lancashire, J.-M. Hou, A. Greystoke, T. H. Ward, R. Ferraldeschi, A. Hughes, G. Clack, M. Ranson, C. Dive, and F. H. Blackhall. Evaluation and Prognostic Significance of Circulating Tumor Cells in Patients with Non-Small-Cell Lung Cancer. *Journal of Clinical Oncology*, 29(12):1556–1563, 2011.

[16] J. R. Landis and G. G. Koch. The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1):159–174, 1977.

[17] C. J. Larson, J. G. Moreno, K. J. Pienta, S. Gross, M. I. Repollet, S. M. O'hara, T. Russell, and L. W. Terstappen. Apoptosis of Circulating Tumor Cells in Prostate Cancer Patients. *Cytometry Part A*, 62A(1):46–53, 2004.

[18] S. T. Ligthart, F.-C. Bidard, C. Decraene, T. Bachelot, S. Delaloge, E. Brain, M. Campone, P. Viens, J.-Y. Pierga, and L. W. Terstappen. Unbiased Quantitative Assessment of HER-2 Expression of Circulating Tumor Cells in Patients with Metastatic and Non-Metastatic Breast Cancer. *Annals of Oncology*, 24(5):1231–1238, 2013.

[19] A. Nanou, F. A. Coumans, G. van Dalum, L. L. Zeune, D. Dolling, W. Onstenk, M. Crespo, M. Sousa Fontes, P. Rescigno, G. Fowler, P. Flohr, C. Brune, S. Sleijfer, J. S. de Bono, and L. W. Terstappen. Circulating Tumor Cells, Tumor-Derived Extracellular Vesicles and Plasma Cytokeratins in Castration-Resistant Prostate Cancer Patients. *Oncotarget*, 9(27):19283–19293, 2018.

[20] J. F. Swennenhuis, A. G. Tibbe, R. Levink, R. C. Sipkema, and L. W. Terstappen. Characterization of Circulating Tumor Cells by Fluorescence In Situ Hybridization. *Cytometry Part A*, 75A(6):520–527, 2009.

[21] M. Tamminga, S. de Wit, J. F. Swennenhuis, E. Heitzer, M. Speicher, T. J. N. Hiltermann, E. M. Schuuring, L. W. Terstappen, and H. J. Groen. Circulating Tumor Cells Measured in the Pulmonary Vein and Radial Artery During Surgery of Non-Small Cell Lung Cancer. In

Proceedings of the American Association for Cancer Research, volume 78, page 5594, jul 2018.

[22] A. G. Tibbe, M. Craig Miller, and L. W. Terstappen. Statistical Considerations for Enumeration of Circulating Tumor Cells. *Cytometry Part A*, 71A(3):154–162, 2007.

[23] G. van Dalum, G.-J. Stam, L. F. Scholten, W. J. Mastboom, I. Vermes, A. G. Tibbe, M. R. de Groot, and L. W. Terstappen. Importance of Circulating Tumor Cells in Newly Diagnosed Colorectal Cancer. *International Journal of Oncology*, 46(3):1361–1368, 2015.

[24] L. L. Zeune. Deep Learning to Identify Circulating Tumor Cells by ACCEPT, 2018.

[25] L. L. Zeune, G. van Dalum, C. Decraene, C. Proudhon, T. Fehm, H. Neubauer, B. Rack, M. Alunni-Fabbroni, L. W. Terstappen, S. A. van Gils, and C. Brune. Quantifying HER-2 Expression on Circulating Tumor Cells by ACCEPT. *PLOS ONE*, 12(10):e0186562, oct 2017.

[26] L. L. Zeune, G. van Dalum, L. W. Terstappen, S. A. van Gils, and C. Brune. Multiscale Segmentation via Bregman Distances and Nonlinear Spectral Analysis. *SIAM Journal on Imaging Sciences*, 10(1):111–146, 2017.

Chapter 6

Deep Learning of Circulating Tumor Cells

Leonie L. Zeune, Yoeri E. Boink, Guus van Dalum, Afroditi Nanou,
Sanne de Wit, Kiki C. Andree, Joost F. Swennenhuis, Stephan A. van Gils,
Christoph Brune and Leon W.M.M. Terstappen

Submitted for publication.

Abstract

Circulating tumor cells (CTCs) found in the blood of cancer patients are a promising biomarker in precision medicine. However, their use is currently hindered by their low frequency, tedious manual scoring and extensive cell heterogeneities. Here, we combined autoencoding convolutional neural networks with advanced visualization techniques to unravel the raw image data of fluorescent images of blood samples enriched for CTCs. We generate a manually scored ground truth set obtained from blood samples of cancer patients that contains over 40,000 images of single cells. Our network classifies cells in five different classes with an accuracy, sensitivity and specificity of over 95% and the obtained CTC counts better predict overall survival of cancer patients than the state-of-the-art counts. Moreover, our network excelled in identifying different subclasses of objects. Deep learning (DL) was faster and superior to classical image analysis approaches and enabled the identification of new biological phenomena.

6.1 Introduction

With the aging of the population, cancer is becoming the leading cause of death. The majority of cancer patients will die from metastasis attributed to the shedding of cancer cells that enter the bloodstream. The peripheral blood load of these Circulating Tumor Cells (CTCs) is strongly associated with cancer prognosis. Their elimination after a first cycle of therapy indicates an effective therapy whereas their continued presence indicates an ineffective therapy [20, 5, 9, 48, 6].

The main achievement of this work is a unified, reliable deep learning framework to identify CTCs in fluorescence microscopy imaging. By automated quantification, identification and visual exploitation of CTCs we move a huge step forward towards automatized analysis and a verified understanding of CTCs in cancer diseases.

Deep learning in healthcare and biomedical imaging. In recent years, the healthcare sector has rapidly moved towards personalized medicine via artificial intelligence (AI), robotics, big data analytics and health informatics [41]. Particularly remarkable is the achievement of AI to overcome the threshold of human analytical performance for an increasing amount of areas in health. Automated data- and machine-driven diagnosis, prognosis and prediction of diseases have the potential to lift healthcare to a new level of quality and understanding overall. In this development, especially radiomics [27] is quickly gaining importance in cancer research. Radiomics describes the high-throughput automated extraction and analysis of quantitative image features from medical images to improve diagnostics and prognosis of disease progression.

Very recently, equipped with the success from computer vision and language processing, advanced machine learning techniques based on deep learning (DL) [29] have started to revolutionize biomedical imaging [30, 3], particularly for large-scale imaging (organ level) in radiology [19, 53]. Besides radiology, most of the deep learning research and advances in cancer analysis can be found in histopathology based on tissue analysis for cancer diseases. Due to an easier access to training data via biopsies, this field has strongly profited from deep learning techniques [43, 45, 23, 50, 31]. Moreover, there have been some first developments on single cell detection [55, 62, 56] and single cell classification or profiling methods [42, 54, 12, 1] via deep learning, although quite general without specific focus towards cancer or CTCs. The main challenges and limitations we specifically observe and tackle within the field of CTCs such as multi-scale, hidden classes, multi-class, dimensionality reduction, uncertainty visualization are not addressed in these papers.

In line with the core idea of radiomics, analyzing CTCs in blood via fluorescence microscopy for cancer diagnosis and monitoring is extremely promising. CTC analysis exploiting sophisticated image analysis methods and the rapid development and validation of medical imaging data that uses image-based signatures for precision diagnosis and treatment, would make it a promising tool for modern biomedicine in clinical practice. However, at the moment, CTCs in the FDA cleared CellSearch® system are identified by *manual reviewing* of thumbnail fluorescent images: cells are required to contain DAPI to identify a nucleus; cytokeratins present in the majority of cells derived from epithelial cancers and CD45

present on the cell surface of leukocytes [11]. An accurate identification and classification of objects as CTCs is essential when CTC counts are used as a basis for disease management of cancer patients. However, for a reliable and reproducible CTC count a clear consensus of experts on an image-based CTC definition is necessary, yet this is currently missing. The current approach by manual reviewing is associated with errors and the introduction of automated image analysis tools can greatly help to improve the reliability of the CTC count [22, 25, 60, 59].

Limitations in current CTC analysis. Identifying CTCs in blood is like finding needles in a complex haystack. There are special tools needed to address the cell heterogeneity in blood samples. Our team has introduced an open source image analysis program, called ACCEPT, which can be downloaded from *https://github.com/LeonieZ/ACCEPT*. ACCEPT enables the analysis, feature extraction and classification of all objects present in full blood sample images (4×175 8-bit images per sample). We have demonstrated that additional, clinically relevant information, such as the presence of tumor derived extracellular vesicles (tdEVs), can be easily obtained with ACCEPT [60, 61, 37, 10]. Yet, our experience in semi-automated CTC analysis has shown us that there is a strong need for a fully automated and robust method for CTC enumeration. An understandable deep learning procedure to identify CTCs in a fast, precise and verifiable manner could answer this need and pave the way towards a more transparent CTC definition for clinical use.

The research context of machine learning methods for tumor cell analysis is still quite unexplored. In particular, a deep learning framework to analyze CTCs for clinically relevant cancer diagnosis and prognosis does not exist yet. Related papers apply *traditional machine learning methods* purely based on extracted features [24, 40, 28, 46, 34, 47]. Although such methods seem to be more transparent by construction, the *bias of user assumptions* in the feature selection limits their performance, particularly in heterogeneous cell regimes. To the best of our knowledge, there exist only two other papers with deep learning concepts for tumor cell analysis. In [2] Chen et al. applied a standard deep learning model for label-free CTC detection by optical phase imaging. Contrary to their label-free imaging approach, we are interested in fluorescence imaging to map our results to a clinically relevant imaging setup like the CellSearch® system. In [35] Mao et al. applied a basic deep learning model for only binary cell classification in the very idealized setup of cell line studies, such that the clinical usability is questionable.

Understandable multiclass cell deep learning. Deep learning in medical imaging is often based on binary classification, e.g. image with disease versus image with control, or image with tumor cell versus image with another cell. Particularly, for reliable analysis of CTCs in the heterogeneous cell composition of the blood, this limitation raises a serious problem because CTCs share feature similarities with other cell types. Moreover, it was shown that also other cell classes can be used as a prognostic biomarker but are currently not exploited. Therefore, there is need for a *systematic multiclass procedure* which can automatically identify all relevant cell classes.

Although many DL methods outperform classical approaches in benchmarks across many different scientific fields, there is a strong demand for more understandable AI in biomedicine [3]. Especially, due to the depth of DL architectures and necessarily large training databases, these models are currently often perceived as black boxes. Researchers in mathematics and machine learning [51, 36] work on general solutions to this problem. Architecture layer visualization [58] and high-level data clustering and visualization [49, 14] is extremely helpful to better understand what happens within DL architectures in critical and discriminating cases. In view of dimensionality reduction, data clustering and robustness, so-called DL AutoEncoders show great potential for medicine. However, only user-friendly visualization tools can allow biomedical experts to exploit hidden structures within data themselves. Thus, to increase fidelity and understanding of DL for cancer research towards clinical practice, there is a strong need for the

- detection of multiple hidden cell classes,
- the inter- and intra- cell class understanding, and
- the visualization of uncertainty due to measurements and limited training data.

Novelty of this work. We explore the use of deep convolutional neural networks and advanced dimensionality reduction architectures to identify and classify objects in fluorescent images obtained with the CellSearch system of patients with metastatic cancer and benign diseases in a clinically relevant setup. We can automatically identify and characterize CTCs in a heterogeneous blood stream in an accurate and reproducible manner not possible before. Our work can be summarized by the following research highlights:

1. *Framework*: For the first time, we propose an automated, accurate and robust deep learning framework for CTC identification, see Figure 6.1, which tackles the heterogeneity and multiclass properties in liquid biopsy.
2. *DL technology*: For the first time, a comparison of Convolutional Neural Network (CNN) classification and AutoEncoder-CNN classification based on latent space analysis has been performed for cells.
3. *Understanding*: For the first time in cancer cell analysis, we achieve understandable AI via an incremental multiclass cell characterization procedure and visualization, which unravels novel subclasses of cells.

6.2 Methods

6.2.1 Generation of the Ground Truth Set

The deep learning network was trained to classify single cells based on an $80 \times 80 \times 3$ fluorescent thumbnail image with three channels. All thumbnail images are extracted from fluorescent images generated by the CellSearch system® (Menarini Silicon Biosystems, Huntingdon Valley PA, USA). After immunomagnetic enrichment of 7.5 ml of blood targeting EpCAM, the enriched cell suspension is incubated with phycoerythrin (PE) conjugated antibodies

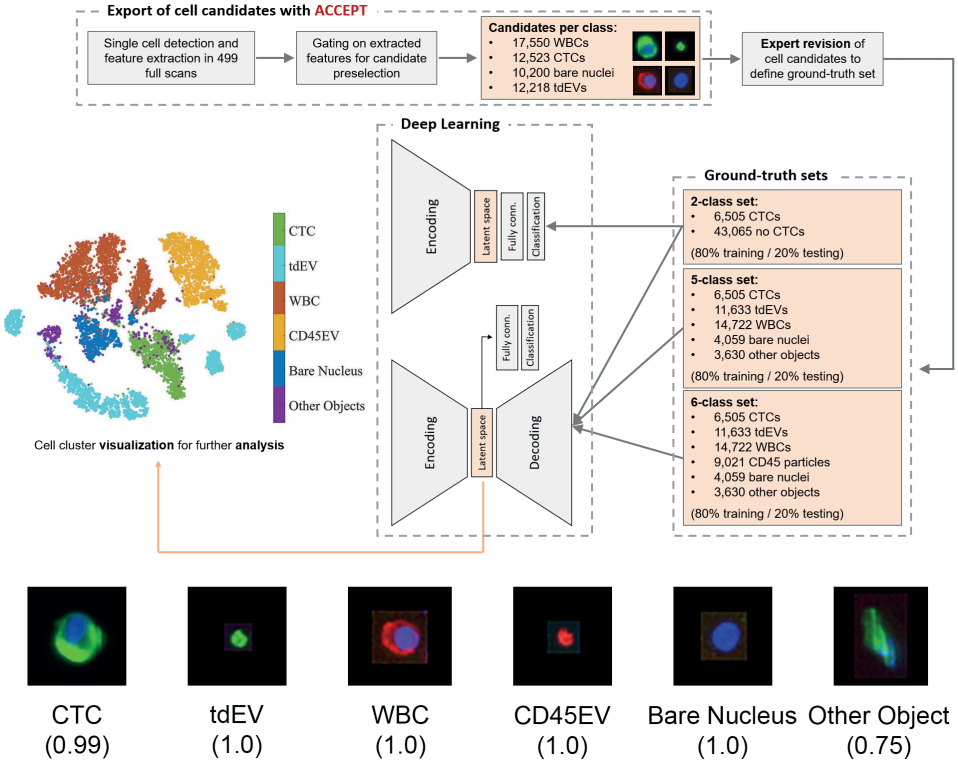


Figure 6.1: *Overview of the analysis workflow used.* Archived images of samples analyzed by CellSearch are first processed with the ACCEPT toolbox to extract thumbnail images of preselected cell candidates. These candidate images are manually reviewed and form the ground truth set for training and validation. Two networks of different architectures are trained and the extracted features in the latent space are used for visualization and further data analysis. An example image for each class is shown at the bottom, together with the probability the network assigned.

directed against epithelial cell specific cytokeratins (CKs), allophycocyanin conjugated (APC) antibody directed against leukocyte specific CD45 and the nuclear dye 4,6-diamidino-2-phenylindole (DAPI) to identify the nucleus. The cell suspension is contained within a cartridge and digital images of the surface for each of the fluorophores are recorded with the CellTracks Analyzer II® (Menarini Silicon Biosystems Inc). The system uses a 12-bit camera and images are transformed and stored as 8-bit images during archiving. Per sample, 175 images for each fluorochrome are generated showing on average around 20,000 cells of different cell classes [10]. All CellSearch images used in this study were obtained from previously reported studies.

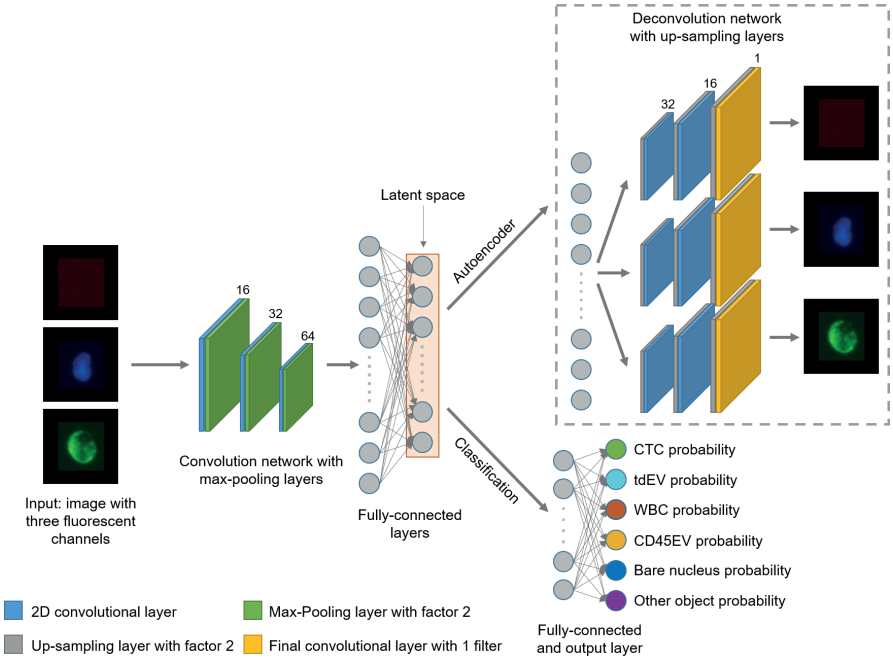


Figure 6.2: Visualization of the convolutional neural network architecture used. The encoding architecture up to the latent space is the same for the standard CNN and the autoencoded CNN. For the standard CNN this is followed only by the classification layers, for the autoencoded CNN a decoding network (gray box) is added and trained simultaneously with the classification layers.

Selection of Samples For the generation of the ground truth set we used in total 499 samples processed with the CellSearch system originating from metastatic breast, prostate, colon and lung cancer patients, patients with benign breast tumors and from eight cell lines. The cell lines used were the prostate cancer cell lines 22Rv1, LNCaP, VCaP, PC3 and DU145 and the breast carcinoma cell lines SKBR-3, MDA-MB 453 and MDA-MB 231. They were all obtained from the American Type Culture Collection (ATCC, Manassa, VA, USA). To efficiently gather a ground truth set with enough CTCs, we mainly used cartridges with high reported CTC counts. We used 199 samples from metastatic breast cancer patients – 99 patients were enrolled in the Detect III study [15] and 100 were enrolled in the BEVERLY02 study [38, 39]. We included 172 samples from metastatic colorectal cancer patients enrolled in the IMMC-06 colorectal cancer study [48]. We used 76 samples from metastatic castration-resistant prostate cancer patients, 26 enrolled in the IMMC38 study [9], 39 samples from a study on androgen receptor expression of CTCs [7] and 11 samples from patients enrolled in the original COU-AA-301 study [16]. We used 39 samples from metastatic non-small cell lung cancer patient originating from studies at the University Medical Center Groningen [11, 10] and 5 samples from patients with benign breast tumors enrolled in the IMMC-26 study [17]. All patients provided written informed consent and more details can be found in the respective listed publications.

Labeling of Cells To generate a ground truth set from the samples exported from the CellSearch system, we first processed all samples with the ACCEPT toolbox for automated CTC analysis. This is an open-source toolbox that can be used to detect all cells and objects in the fluorescent images by advanced multiscale segmentation [61, 60]. Moreover, for each cell and each fluorescent channel several measurements like mean intensity, size and roundness are extracted and can be used to identify, classify and characterize all objects by the application of linear gates [37]. To find possible candidates for each of the four original cell categories (CTC, tdEV, WBC, bare nuclei), we applied very loose linear gates on the extracted measurements and all candidates falling into one of these gates, were extracted. The gate settings used were:

	PE Mean Intensity	DAPI Mean Intensity	CD45 Mean Intensity
CTC	> 0	> 0	≤ 0
tdEV	> 0	≤ 0	≤ 0
WBC	≤ 0	> 0	> 0
Bare Nucleus	≤ 0	> 0	≤ 0

The fluorescent thumbnail images of the cells falling in the gates were saved as tiff images using ACCEPT and were sorted based on the category they fell into. For each of the four groups we generated random sets of 150 thumbnail images and these sets were reloaded into the ACCEPT toolbox to manually score if the extracted candidate really belongs to the respective cell class. All images are presented in a thumbnail gallery where the signal in each fluorescent signal is shown individually and an overlay of the CD45, DAPI and PE fluorescent channels next to it (see Supplementary Figure 6.2). By clicking on the overlay image a cell can be labeled and unlabeled. By default, all events were labeled as being a CTC, tdEV, WBC or bare nucleus and reviewers had to go through the thumbnail images to unlabel the ones which did not fall into the class by clicking on the corresponding overlay image. This work was done by five different reviewers experienced in manually scoring CTCs and other cell types. The overlay images that were not labeled afterwards and thereby did not show a cell that fell into one of the four cell classes, were reviewed again to find examples of artifacts and objects of unknown origin. To enlarge the set of images showing fluorescent artifacts or objects of unknown origin, we added examples of background artifacts exported from the five samples of patients with benign breast tumors. These samples appeared to have a very high fluorescent background and many thumbnail images in ACCEPT show background artifacts that were mistakenly segmented. Although not showing any cells, these thumbnail images have to be classified by the DL network as soon as it is applied to full sample scans by detecting every object with the ACCEPT toolbox and then classifying it with the deep learning network. All objects of unknown origin were summarized in a fifth class and together with the four cell categories they formed the 5-class set. Before it was used for training, the ground truth set was double-checked by yet another reviewer to reduce the noise in the labels which is inevitable when manually scoring a huge amount of cells.

Image Preprocessing Before the thumbnail images in the 5-class set were used for the training of the network, some preprocessing steps were applied. All fluorescent channels apart from the first three used to identify CTCs, i.e. DAPI (nucleus), CK-PE (cytokeratins), CD45-APC (leukocytes), were removed. In the ACCEPT toolbox, the interpretation of the channels is set: DAPI is an inclusion marker as the presence of a nucleus is required, and CK-PE is an inclusion marker as cytokeratins are expressed in the majority of cancer cells of epithelial cell origin, and CD45-APC is an exclusion marker present on WBC that should not be present on CTCs. The interpretation of the other fluorescent channels is frequently used to assess the expression of certain antigens on the CTCs and not for their identification and therefore these channels were not used in our network. All thumbnail images need to have the same size which is large enough to cover all single cells but also artifacts. Moreover, the size in x- and y-dimension is reduced by a factor of eight during the forward pass through the network and, in case of the autoencoding framework, it is afterwards increased with the same factor. To avoid any influence from rounding image sizes to the next integer, we decided for an image input size of $80\times80\times3$ which is divisible by eight in x- and y-dimension. We resized the images to that format by adding a symmetric zero padding in both x- and y- dimension if the original thumbnail size was smaller than 80 so that the object to be classified is located in the middle of the thumbnail image. Moreover, the 12-bit images were cast to 8-bit images and the three fluorescent channels were merged and saved as one RGB color image which was input to our network.

6.2.2 Deep Learning Network Architectures

To automatically classify the fluorescent thumbnail images of cells into separate cell classes, we used convolutional neural networks [29, 26]. We used two different architectures that are presented in Figure 6.2. The first architecture is a standard CNN with three convolution layers with a kernel size of 3×3 and an increasing number of filters. The first convolutional layer has 16 filters, the two following ones have 32 and 64 filters. Each convolutional layer is followed by a max-pooling layer to halve the resolution of the input. Thus, the input image is transformed from a $80\times80\times3$ image to a $10\times10\times64$ representation of the input by forward-passing it through the three blocks of convolution and max-pooling layer. The convolutional part is followed by a fully-connected layer of 208 neurons which are connected by another fully-connected layer to 50 neurons that form the *latent space*. In this layer, no activation function is applied. This 50-dimensional latent space representation is input to the final two fully connected output layers with six neurons each, where the last one has a softmax activation function so that the last six outputs sum up to one and can be interpreted as the probability of belonging to each class. To train this network, we minimized the cross-entropy loss function

$$L_{\text{CNN}}\left(y_{\text{true}}, y_{\text{pred}}\right) = -\sum_i y_{\text{true}}(i) \log\left(y_{\text{pred}}(i)\right),$$

where y_{true} is the one-hot encoded true label for a given input x_{inp} and y_{pred} is the vector of predicted class probabilities. For the second network architecture, we used convolutional autoencoders to enhance the results [13, 18]. Autoencoders first encode the input data and

then reconstruct the input signal from the encoded representation as good as possible. In our case, the latent space containing the encoded input data has a much lower dimension than the input space. In that way, the network learns a lower-dimensional representation that still contains most of the variation of the input data in order to reconstruct the input from the encoded latent variable. For classification tasks, autoencoders are often used to pre-train networks especially if the amount of labeled ground truth examples is too small to optimize very deep networks. In this case, the weights are first optimized on unlabeled data and in a second step the network is fine-tuned to learn the classification based on labeled input data [13]. Contrary to this approach, we used the autoencoder and the classification part simultaneously [18, 32]. The architecture for encoding and classifying is the same as for the standard CNN used, but the latent space representation is now also used as an input to a decoding convolutional network (gray box in Figure 6.2). The decoding network starts with a fully-connected layer of 6400 neurons whose activation is reshaped to a tensor of size $10\times10\times64$, which is input to the convolutional part of the decoding network. In our tests, we saw that we obtain superior results if we have a separate convolutional decoding network for each of the three fluorescent channels. Each network contains three blocks consisting of an up-sampling layer with factor two followed by a convolutional layer with kernels of size 3×3. For all three networks, the first convolutional layer has 32 filters, the second one 16 filters and the final layer has one filter. The output layers use a sigmoid activation function and all other layers whose activation is not specified so far use rectified linear unit (ReLU) activations. To train the autoencoded classification network, we minimized the following loss function:

$$L_{\text{AE-CNN}}\left(\left(x_{\text{inp}}, y_{\text{true}}\right), \left(x_{\text{rec}}, y_{\text{pred}}\right)\right) = \sum_{c=1}^{3}\left(-\sum_{i,j} x_{\text{inp}}(i,j,c)\log\left(x_{\text{rec}}(i,j,c)\right)\right) + \beta \cdot L_{\text{CNN}}\left(y_{\text{true}}, y_{\text{pred}}\right).$$

Here, x_{rec} is the reconstruction of the input signal based on the compressed representation in the latent space. During training we noticed that the algorithm often converged to a local minimum of weights that resulted in very high activation in the latent space and in the reconstruction all values were close to zero. Therefore, we added an L^2 regularization on the latent space activation which prevents the algorithm from converging to this weight-set.

Training Parameters and Setup We implemented and trained the networks using Keras [4]. To enlarge the number of example images for each cell class, we used the build-in image augmentation function. We divided the input image by 255 (thus the intensity was scaled from 0 to 1), and allowed: random rotations up to 40°; random shifts in x- and y-direction up to 15% of the total image range; random zoom between 80% and 120% and random horizontal flips. Moreover, we applied a customized background subtraction method where we subtract in each fluorescent channel the lowest 10% of the intensity values present. This accounts for the high fluorescent background in some thumbnail images. We minimized the loss function with the Adadelta algorithm [57] in 75 epochs. To increase the stability of the learning process, we doubled the batch size every 15 epochs starting with a batch of 16

images. For training the autoencoded CNN, we used an iterative training scheme: We first trained the network with $\beta = 0$, so that only the autoencoder was trained and the weights in the classification layers stayed random. The resulting weights were used as an initialization for another training phase with $\beta = 10$. Then we used these weights again as an initialization and increased β to 100 and in the last phase, we trained the network with $\beta = 1000$. Thus, in every iteration we increased the influence of the classification performance on the weights. We evaluated the resulting networks in terms of accuracy, sensitivity, specificity and performance on patient samples and used the third network, trained with $\beta = 100$ as the final network to generate the results.

6.3 Results

6.3.1 Composition of Training and Validation Sets

The original data set used to train and validate our networks was obtained through the automated processing of 499 patient samples with ACCEPT and contained fluorescent images of 6,505 circulating tumor cells (CTCs), 11,633 tumor-derived extracellular vesicles (tdEVs), 14,722 white blood cells (WBCs), 4,059 bare nuclei and 3,630 other objects (Figure 6.1). During the training of the network we detected a class of objects that occupied a unique space in the encoded data representation and was characterized by the expression of CD45 and the lack of DNA and cytokeratin (see section “Identification of New Classes” below). As these objects lack a nucleus and have similar morphological characteristics in terms of shape and size as the tdEVs, with the only difference that they express CD45 instead of cytokeratin, we baptized them CD45 Extracellular Vesicles (CD45EVs) and we added 9,021 fluorescent images of these CD45EVs to the total ground truth set. Typical example images of the six classes are shown at the bottom of Figure 6.1. Throughout our experiments we used this data set in three different compositions:

1. **2-class Set:** The data set was divided into two classes –“CTC” and “no CTC” class.
2. **5-class Set:** The data set was divided into five classes – “CTC”, “tdEV”, “WBC”, “Bare Nucleus” and “Other object”.
3. **6-class Set:** The data set was divided into six classes – “CTC”, “tdEV”, “WBC”, “CD45EVs", “Bare Nucleus” and “Other object”. The 9,021 images from the CD45EVs were added to the 5-class set.

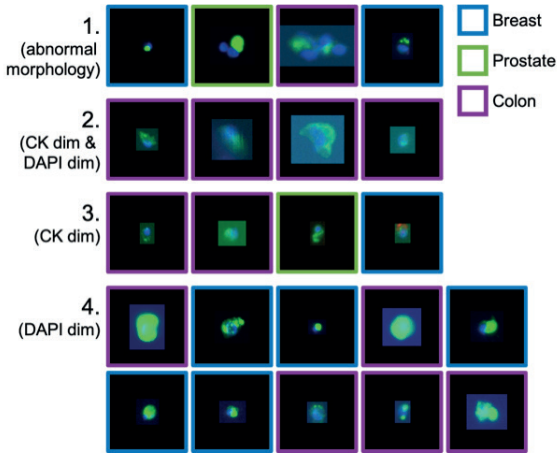
Each set was randomly split into a training set of 80% and a validation set of 20% of the images.

6.3.2 Automated Classification of Circulating Tumor Cells

We trained a convolutional neural network (CNN) for the 2-class classification of fluorescent images of CTCs versus thumbnail images of all other objects found in the fluorescent images. The classification performance on our validation set of 1,301 CTCs and 8,618 other objects is

Accuracy: 98.88 %		Predicted Class: CTC	Predicted Class: no CTC
True Class	CTC	1279	22
True Class	no CTC	89	8529

(a) Confusion Matrix.



(b) Misclassified CTCs.

Figure 6.3: Classification results of the 2-class CTC classification network. Panel (a) shows the confusion matrix after classifying the validation set. The matrix shows the absolute number of cells for each category. Panel (b) shows thumbnail images of the 22 misclassified CTCs. For visualization we scaled each channel from the minimal to maximal value present in the thumbnail. The cells can be grouped into 1. cells with abnormal morphology, 2. cells with dim CK and dim DAPI staining, 3. cells with dim CK staining and 4. cells with dim DAPI staining. The color border of each thumbnail indicates the tissue of origin.

summarized in 6.3 (a). Overall, we received an accuracy of 98.9% with a sensitivity of 98.3% and a specificity of 99.0%. The network assigns to each object a probability for belonging to a class and for the 1,279 correctly classified CTCs only 91 (7.1%) had a probability of being a CTC lower than 90% out of which 25 (1.9%) had a probability less than 75%. For the 89 objects classified as a CTC although they are not, 67 had a CTC probability higher than 75% and out of these, 50 had a probability higher than 90%. The 22 CTCs misclassified by the network are shown in Figure 6.3 (b).
Two of them originate from prostate cancer patients (marked in green), eight from breast (marked in blue) and 12 from colon cancer patients (marked in purple). This results in 0.8% of all prostate cancer (PC) patient CTCs, 1.2% of all breast cancer (BC) patient CTCs and 4.2% of all colon cancer (CC) patient CTCs being misclassified. All cells from cancer cell lines (CCL) and CTCs from lung cancer (LC) patients in the validation set were correctly classified. The lung cancer patient CTCs represent however only 1.6% of all CTCs tested. The misclassified CTCs can be grouped into four categories based on their staining and morphology as illustrated in Figure 6.3 (b).

The CNN we used can be divided into two parts, a convolutional feature extraction part (called “Encoding” in Figure 6.1) followed by the fully-connected layers classifying the input based on the features extracted in the first part, see Figure 6.2. We refer to the lower dimensional space after feature extraction as the *latent space*. In our case the inputs of

size $80 \times 80 \times 3$ are compressed to a 50×1 feature representation. Similar inputs result in a similar feature representation and are therefore close in the latent space. To visualize the latent space, we compressed the 50-dimensional feature space to a two-dimensional space using t-distributed stochastic neighbor embedding (t-SNE) [49]. t-SNE is a nonlinear dimensionality reduction method that maps similar input vectors to close points in the reduced space and keeps the distance between clusters from the original, higher-dimensional space. Figure 6.4 (a) shows the two-dimensional t-SNE map of all CTCs in the validation set after feature extraction. All points are colored based on their cancer type. In the plot we highlighted three clusters by black boxes; two on the opposite sites of the point cloud and one in the middle. For all three groups, we randomly chose five different points and visualized the corresponding thumbnail images in panel (b). As illustrated by the three examples, the location in the latent space reflects typical morphological and staining features. The cells in the lower-right corner of the latent representation (group 3) share a relatively large cell size, strong CK staining (green) and round morphology. Cells in the middle cluster (group 2) are smaller, the morphology is less circular and the CK staining not as bright as in group 3. The cells in the top-left corner (group 1) have either a very dim staining or an abnormal morphology. They strongly overlap with the set of misclassified CTCs shown in Figure 6.3 (b). We further analyzed if the CTCs of different cancer types have specific morphological features separating them from the other cancer types. Yet, it can be seen in the frequency plots of the latent space in panel (c) of Figure 6.4 that the latent space representations of CTCs do not cluster per cancer type. In the frequency plots, we plotted for each location the number of CTCs of that cancer type that can be found in a 5×5 window around that point. CTCs from breast, prostate and colon cancer have a very similar distribution reflecting the shape of the whole point cloud. The amount of CTCs from lung cancer patients and cells originating from cancer cell lines is much lower which most probably explains the more restricted distribution across the point cloud.

6.3.3 Detection of Cell Classes by Autoencoded Deep Learning

Label-independent clustering of cell classes Apart from the standard classification network, we also used a network architecture where the encoding network is followed not only by the fully-connect classification layers, but also by a decoding network (Figure 6.1). As shown in Figure 6.2, the latent space representation is input to the classification but also to a deconvolution network (gray box) that maps the input back to the image space. To investigate the structural difference of the latent space for the standard and the autoencoding convolutional classification network, we trained both networks on the 2-class set of CTCs and no CTCs. Figure 6.5 shows the two-dimensional t-SNE maps of the latent space. While both algorithms map the input images to different clusters based on the 2-class label, we see that the structure of the red class is significantly different for both network architectures. While the standard network uses the latent space representation only to distinguish between CTCs and no CTCs, the autoencoding network needs to reconstruct the $80 \times 80 \times 3$ input image as good as possible only based on the 50-dimensional feature vector in the latent space. Therefore, the latent space shows distinct clusters of objects that look similar in

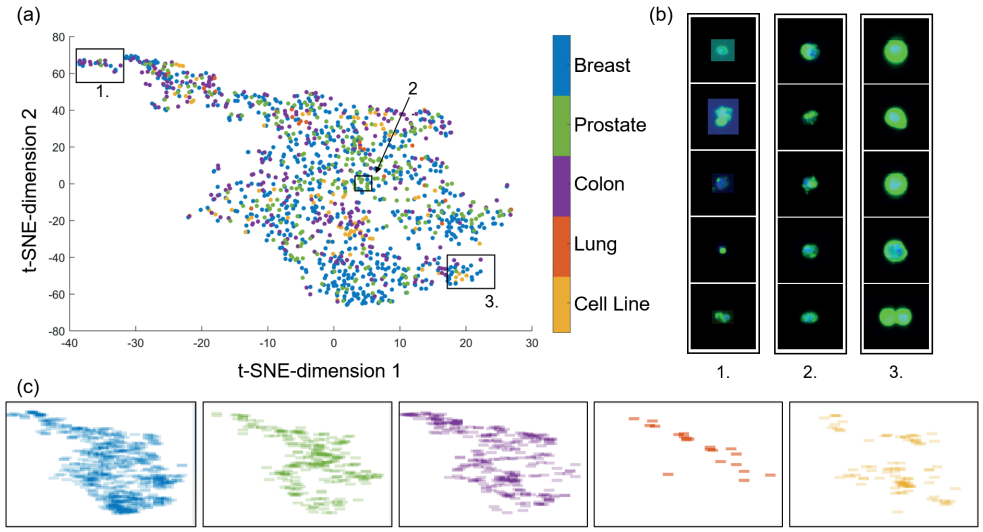
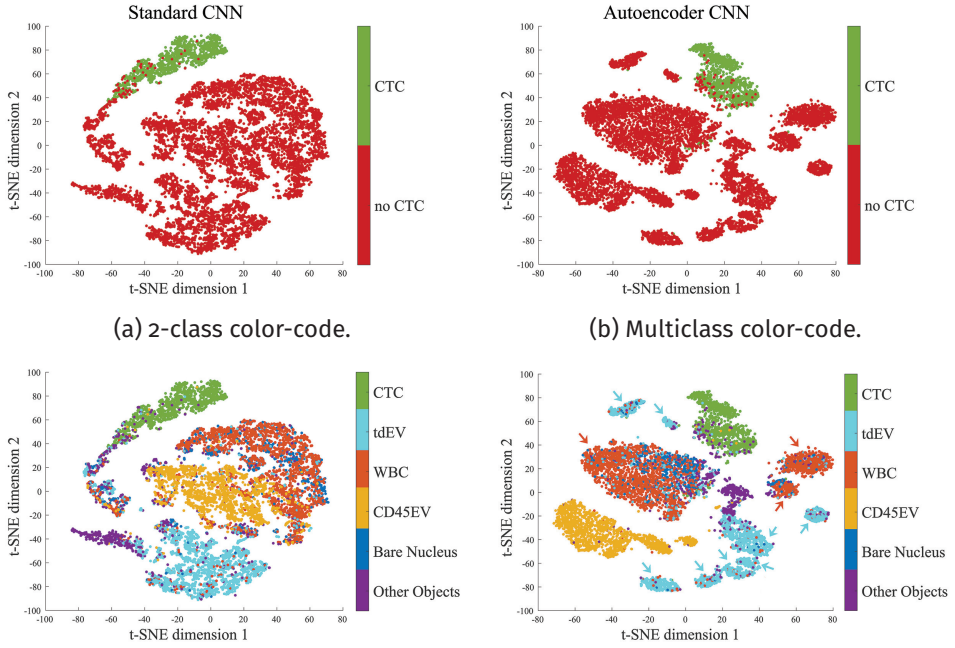


Figure 6.4: Two-dimensional t-SNE map of all CTCs in the validation set colored based on the cancer-type (panel (a)). In panel (b) we show five cells that we randomly chose from each of the three locations highlighted with a black box in the plot (top-left, middle and bottom-right of the point cloud). We see that the position in the latent space reflects morphological features like staining intensity, size and roundness. Panel (c) shows frequency plots for each cancer type.

the original image domain. In Figure 6.5 (c) and 6.5 (d) we colored the data points based on the true labels from the six original classes (CTC, tdEV, WBC, CD45EVs, bare nuclei and other objects). Although both networks were not trained to distinguish all six classes, we see that the cell classes cluster. The autoencoding architecture on the right has the advantage of very distinct clusters. While in the standard architecture on the left, the bare nuclei are distributed across the WBC cluster with no obvious pattern, in the autoencoding architecture a transition of the bare nuclei to the top right corner of the WBC cluster is observed. This reflects the weak CD45 signal, especially of the granulocytes, recorded by the CellTracksAnalyzer II. The majority of the objects classified as bare nuclei are in fact WBCs [10]. Thus, a shared cluster with a smooth transition between these two cell classes reflects our expectation. Another advantage of the autoencoding network structure is the detection of cell subclasses. In Figure 6.5 (d) light blue arrows indicate eight different clusters of tdEVs. The implication of this finding still has to be investigated by linking the clinical observations with the number of tdEVs in each subclass. The presence of three WBC clusters indicated by the brown arrows likely indicates the presence of WBC subpopulations.

Identification of new classes Using the autoencoding classification network we can detect previously unobserved classes as illustrated in Figure 6.5 (d). The class of CD45EVs was identified as a unique class in such manner as illustrated in Figure 6.6. Different from the



(a) 2-class color-code.

(b) Multiclass color-code.

(c) 2-class color-code.

(d) Multiclass color-code.

Figure 6.5: Comparison of the t-SNE latent space representation for a standard and an autoencoding convolutional classification network. In (a) and (b) points are colored based on the true 2-class label that the network was trained to distinguish, green for “CTC”, red for “no CTC”. In (c) and (d) points are colored based on the true 6-class label (“CTC” green, “tdEV” light blue, “WBC” brown, “CD45EV” orange, “bare nucleus” blue, “other object” purple) which was not used during training.

scatter plots before, the points are now color-coded based on the *predicted class*, not the true class label. The network was trained and validated on the 5-class data set (CTC, tdEVs, WBC, bare nuclei and objects of unknown origin). If we add now the sixth class, CD45EVs, to the validation set, we see a very distinct cluster of objects highlighted by the gray ellipse. Within this cluster we see very diverse predicted classes, in particular mixing class labels for very distinct cell classes like tdEV (light blue) and WBCs (brown). This is an indicator that the network has not seen this type of cell before and is not trained to predict the correct class. Yet, due to the autoencoding architecture the network learns that these cells are different from the other cell classes and form a distinct cluster in the latent space. This type of cluster is a hint that the set that was encoded and classified with the network contains a class of cells that was not in the training set. We randomly chose 15 cells that belong to a data point in the black box on the left site of the point cloud in Figure 6.6. The corresponding fluorescent images are shown on the right site of the graph. As we can see, they all belong to the new class of CD45EVs that was not included in the set when we trained the network. This approach can be used in an iterative procedure to include more and more cell classes

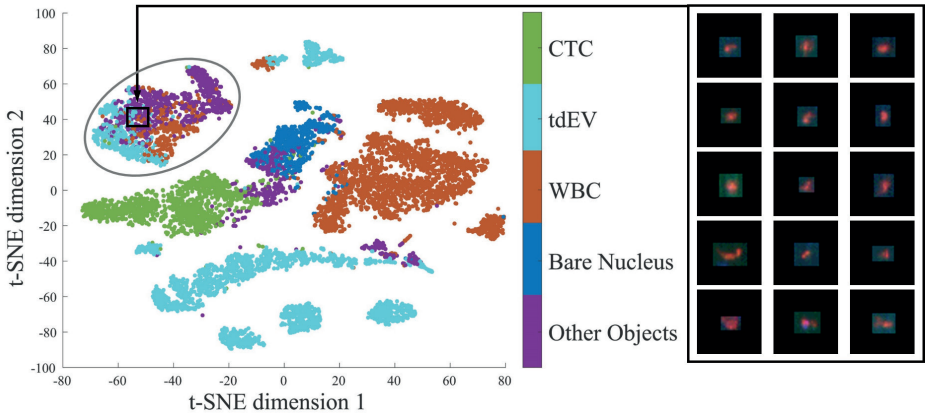
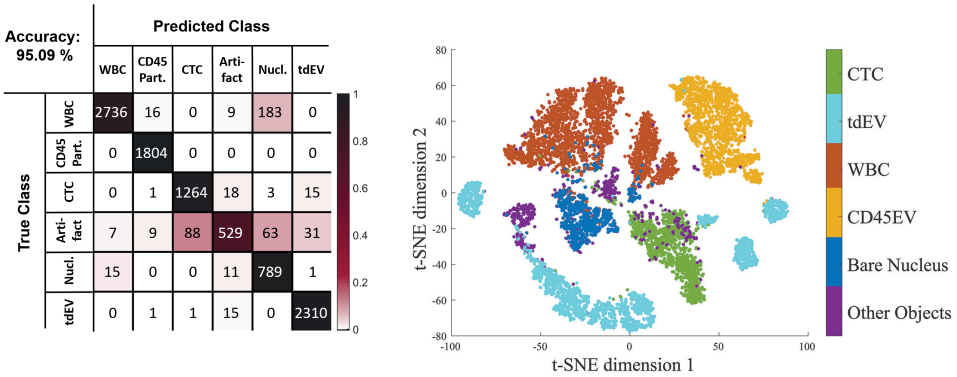


Figure 6.6: Identification of new cell classes based on the two-dimensional t-SNE map. Points are colored based on the *predicted* class of the autoencoding network. A class which is represented by a very distinct cluster yet shows very diverse prediction labels (gray ellipse) indicates a class of cells which is so far unknown to the network. We randomly chose 15 points located in the black box in the t-SNE map and plotted the corresponding cells which are all CD45EVs, a class which the network has not seen during the training phase.

automatically found by the network with the goal of capturing the large variety of cells, cell fragments and artifacts that can be found in the fluorescent images.

6.3.4 Automated Multi-Class Cell Classification

We ultimately trained the autoencoding convolutional classification network to differentiate between all six classes and obtained an overall accuracy of 95.1%. The classification results and the resulting latent space representation of the validation set are summarized in Figure 6.7. The sensitivity for classifying CTCs was 97.2% with a specificity of 99.0%. Thus, the specificity remained the same in comparison to the 2-class classification network described before, while the sensitivity slightly decreased by 1.1%. In total, 37 true CTCs were not recognized by the network as a CTC, out of which 14 originated from breast cancer samples, 19 from colon cancer and 4 from prostate cancer. One out of these 37 CTCs was now classified as a CD45EV, 18 as “other objects”, three as bare nuclei and 15 as tdEVs. Yet, only 17 decisions were made with a class probability of over 75% and if we increase the threshold to 90% this reduces to 8 cells. The 17 misclassified CTCs with a class probability higher than 75 % are shown in the Supplementary Figure 6.1 (a). Analysis of the 89 cells that were classified as a CTC by the network but not by the manual reviewers, showed that, according to the given label, one was a tdEV from a lung cancer sample and all 88 others were in the class of other objects. One originated from a lung cancer sample, 44 from breast cancer, five from samples of cancer cell lines, 18 from colon cancer and 20 from prostate cancer samples. Out of these 89 cells, 45 had a probability of being a CTC over 75% and out of these 17 had a probability



(a) Confusion Matrix. (b) t-SNE map.

Figure 6.7: Classification results and encoded representation of the 6-class validation set in form of a two-dimensional t-SNE map. Panel (a) shows the confusion matrix for the 6-class classification task. The matrix contains the absolute values and is colored based on the relative frequency for each field where the frequencies per row add up to 1. Panel (b) shows the t-SNE map of the latent space representation of the validation set. Data points are colored based on their *true* class label.

higher than 90% (presented in Supplementary Figure 6.1 (b)). If we analyze for each cancer type the amount of cells that were misclassified, we see that 15% of the cells coming from benign breast samples were classified wrong (n=185), 8% of all lung cancer cells (n=892), 6% of both the colon cancer cells (n=1231) and the breast cancer cells (n=3460), 5% of the cells coming from cancer cell lines (n=369) and only 2% of the cells from prostate cancer samples (n=3782). Apart from the objects exported from cancer cell lines, there seems to be an inverse correlation between the number of examples used per cancer type and the percentage of misclassified data points of this cancer type. This encourages the addition of more examples of cells from under-represented cancer types to both the training and the validation set in order to improve their recognition.

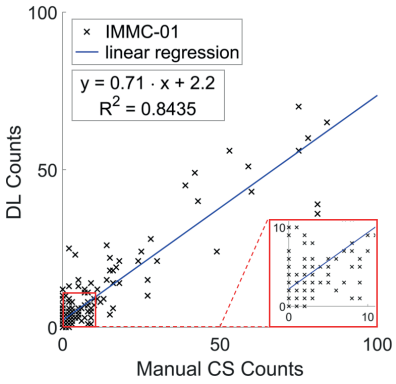
6.3.5 Deep Learning Classification of Full Patient Samples

To test the performance of our DL network on unseen examples, we exported thumbnail images together with their manual score from the CellSearch® system. For all samples, we exported all thumbnail images that were presented to reviewers in the CellSearch® system together with the score given by the reviewer. In total, we exported 250 randomly chosen samples from patients with metastatic breast cancer enrolled in the original IMMC-01 prospective, multi-center trial [8] and used no replicates originating from the same patient. The images from the IMMC-01 study were acquired on CellSpotters (Nikon fluorescence microscopes) and are therefore recorded with a different microscope than the samples included in our training and validation set.

We excluded all samples with more than 10,000 presented thumbnail images and all

	Overall Agreement	Agreement on CS CTCs	Cox HR on IMMC01 Subset
CS	100%	100%	2.4 95% CI (1.4-4.1)
DL without cut-off	94%	53%	4.1 95% CI (2.3-7.4)
DL with 50% cut-off	94%	52%	3.9 95% CI (2.2-6.9)
DL with 75% cut-off	95%	45%	3.3 95% CI (1.9-5.5)
DL with 90% cut-off	95%	33%	4.0 95% CI (2.3-6.8)
DL with 99% cut-off	94%	13%	6.7 95% CI (3.3-13.3)

(a) Comparison of the CTC counts for different cut-offs for the DL probability.

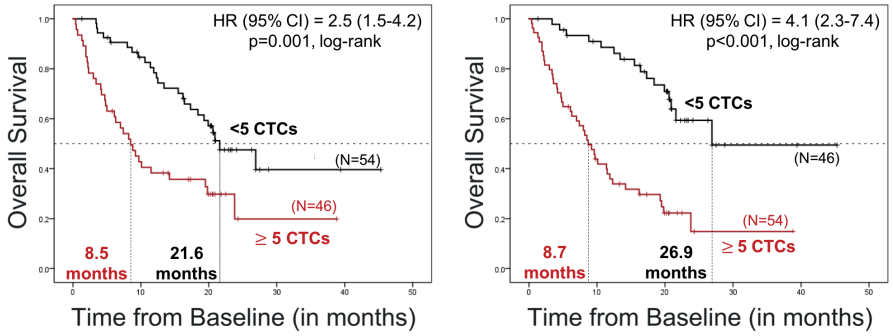


(b) Scatter plot of manual CellSearch CTC counts vs DL counts.

Figure 6.8: *Comparison of the CellSearch CTC counts vs. DL CTC counts.* In table (a) we compare the overall agreement with the manual counts on all 23,377 events if all cells with a probability below the cut-off are counted as no CTC and all above are counted as a CTC. Besides that the agreement with the manual counts only on the 1,526 CTCs is reported. The last column shows the Cox hazard ratio for the survival analysis on the subset of the IMMC-01 patients together with the 95% confidence interval. The scatter plot in panel (b) compares the manual CellSearch CTC counts with the CTC counts obtained from the DL classification for all analyzed IMMC-01 samples. A linear regression is shown in blue and a magnification of the scatter plot around the origin is shown on the right.

samples with 100 or more CTCs in CellSearch®. This was done to prevent inclusion of samples where the CTC count was extrapolated due to high CTC numbers. This resulted in 227 patient samples with a total of 23,377 events to score. Out of all events, 1,526 were manually scored as a CTC by the reviewers and the other 21,851 were rejected. If we classify the same thumbnail images with the DL network, we receive an overall agreement of 93.6%. For the 1,526 CTCs the network agreed in 813 cases (53.3%). If we increase the threshold of the probability for an event to be counted as a CTC, the overall agreement remains fairly constant while the agreement on the CTCs decreases. Figure 6.8 (a) summarizes the results for a threshold of 50%, 75%, 90% and 99%. Figure 6.8 (b) summarizes the number of manual CTCs per samples versus the number of CTCs classified by the DL network (without setting a cut-off for the CTC probability) for all 227 samples. Analysis of cells where the two counting methods disagree shows that the network scores more CTCs with a dim CK staining or apoptotic CTCs. On the other hand, reviewers score more cells with a very bright CK but very dim to nearly absent DAPI staining as a CTC where the network decides for a tdEV. The automated classification of all 23,377 thumbnail images on a computer with a single GPU took less than 3 minutes.

We further investigated whether the number of CTCs obtained by the DL network is correlated with the overall survival of a patient. For better comparison we only used samples drawn before the initiation of a new therapy. Moreover, we excluded samples where the reported CTC count deviated from the count that we exported from the CellSearch® system. In



(a) Kaplan-Meier Plot for CellSearch counts.

(b) Kaplan-Meier Plot for DL counts.

Figure 6.9: *Comparison of the predictive value of the CellSearch scores and DL scores by Kaplan-Meier plots.* The DL based scores (b) give a better separation of the survival curves for patients with favorable and unfavorable prognosis compared to the "ground truth" score in (a). The median overall survival for each group, the Cox hazard ratio (HR) and the significance (log-rank p) are indicated in each panel.

total, we ended up with 100 patients included in our analysis. Figure 6.9 shows the Kaplan-Meier plots for patients with favorable and unfavorable CTC counts for both, the manual scores (Figure 6.9 (a)) and the DL scores (without threshold, Figure 6.9 (b)). Median overall survival for patients with less than 5 manual CTCs was 21.6 months versus 8.5 months for patients with 5 or more manual CTCs ($p < 0.001$, log-rank-test). For the automated CTC counts, patients with less than 5 DL CTCs had a median overall survival of 26.9 months whereas patients with 5 or more DL CTCs had a median overall survival of 8.7 months ($p < 0.001$, log-rank-test). We see that the CTC counts from the DL network result in a better discrimination between patients with 5 or more CTCs and patients with less than 5 CTCs than the manual "ground truth" score. This is also reflected by the higher Cox hazard ratio in Table 6.8 (a). The best discrimination can be obtained by either using no threshold (HR of 4.1) or using a high threshold of 90% (HR of 4.0). The very high threshold of 99% results in a very high Cox hazard ratio but also in unbalanced groups of 86 and 14 patients with favorable and unfavorable prognosis respectively and needs further investigation with a larger cohort. Yet all DL based CTC counts result in a better discrimination than the manual counts (Figure 6.9 (a)). This raises the question whether the manual CellSearch® count really is the "true" CTC count that we aim to achieve.

6.4 Discussion

In the recent years convolutional neural networks and deep learning methods have shown to give remarkably good results in various applications ranging from speech recognition [21]

and classification [26], to image processing tasks like denoising [52] and segmentation [44]. Nevertheless, in order to be used in the biomedical field with clinical relevance, understanding the way a CNN interprets the data is of utmost importance to reduce the uncertainty in the approach. In this paper we introduced an autoencoded multi-class classification network which has, in combination with the proposed visualization and analysis of the latent space, the potential to analyze cell classes and their relations in a new way and explore novel subclasses of cells in a structured manner. We showed that our network accurately classifies CTCs and can be used for a fully-automated CTC analysis framework. Our network is also able to classify other cellular objects in fluorescent images of EpCAM enriched blood samples of cancer patients.

In a previous publication [14] it was shown that t-SNE visualizations of the feature map can help to identify new cell classes. Yet, with a standard deep learning network only trained for classification, the transition between cell classes that were not labeled separately in the ground truth set is often continuous. This can be problematic for new cell classes if they are not significantly different from all other classes in the ground truth set. We have shown that the combination of the classification network together with an autoencoder leads to more distinct clusters in the latent space, even for cell classes not specifically labeled before. This makes the visualization of the latent space after dimensionality reduction with t-SNE even stronger and has the potential to iteratively explore more distinct subclasses in various applications. Another advantage of the autoencoded classification network is its increased robustness towards noise in the ground truth set. With manually scoring several thousand cells, it is unavoidable that some mistakes can be found in the scores; even when the data is double-checked. This is visualized in Supplementary Figure 6.1 (b) where we present cells that are – according to the given label – not CTCs but the majority of them actually are CTCs. The autoencoding network forms very distinct clusters of similar cells in the latent space making a seperation of the cell classes for classification easy. Even if the noisy labels slightly change the decision boundaries between cell classes, this will have only little effect on the final decision. For a standard network, where the representation of cell classes in the latent space tends to be much more continuous, a slight change of the decision boundaries can have a much higher impact on the final outcome. Another question is how to deal with objects that cannot be clearly assigned to a cell class, either because they are artifacts in the fluorescent images or because it is unclear whether they are a cell with an uncommon morphology or not. Although dealing with this question is not needed for a manually created ground truth set of thumbnail images, this question needs to be addressed as soon as we want to classify all objects found in the fluorescent images with the network. We expected that the autoencoding approach would be sufficient to seperate them from the other objects, but noticed that some of them are still represented by a point in the latent space that is very close to a cluster of actual cells, making it hard to seperate them if the network is not trained to. Therefore, we used a class for “other objects” where we present those kind of objects. Due to the nature of that very diverse “rest class”, it is not defined by a clear distinct cluster in the latent space, but can be found in multiple locations. Although this makes it still difficult to reliably separate all of these objects from the other cells, we saw a clear improvement in identifying those events. Moreover, it is questionable if its really necessary to receive a

good classification accuracy on these objects or if it is enough that the extra class prevents that they are classified with a high probability in one of the other real cell classes and would bias the resulting cell counts we obtain.

In a set of 100 metastatic breast cancer patient samples processed with the CellSearch® system, we have shown that the number of events classified by the DL network as a CTC is correlated with the overall survival chance of a patient. The network assigns to every classified event a class probability for each of the six classes. In particular, it assigns a probability of being a CTC to all events. Yet it remains open for discussion whether all events where the probability of being a CTC is the highest class, should be counted as a CTC, or whether a certain CTC probability should be exceeded to be counted. We have shown that for the group of 100 patients, all DL based counts, ranging from counting all CTCs to counting only the ones with a very high probability of over 99%, can be correlated to the patients' outcome. Dichotomization of the patients based on a cut-off of 5 CTCs even leads to a better separation of patients with favorable and unfavorable prognosis than the manual CTC count, considered as the golden standard for CTC counting and seen as the "true" answer. This raises the question whether the manual count really is the best answer and should be used as the reference count that we want to achieve, or if the classification performance of the network should be solely judged by the prognostic value of the obtained cell counts. Nevertheless, to obtain a fully automated method with reproducible counts, a defined, fixed threshold is needed. To analyze what the best choice is, further survival analysis with a larger patient cohort, in particular with different cancer types, is needed. On the one hand, a lower threshold for the CTC probability increases the reported number of CTCs which might be beneficial for cancer types with on average lower CTC counts than in breast cancer patients. On the other hand, there is a quest to reduce the current cut-off value of 5 CTCs to only 1 CTC which could be based on a stricter definition of a CTC obtained by increasing the threshold.

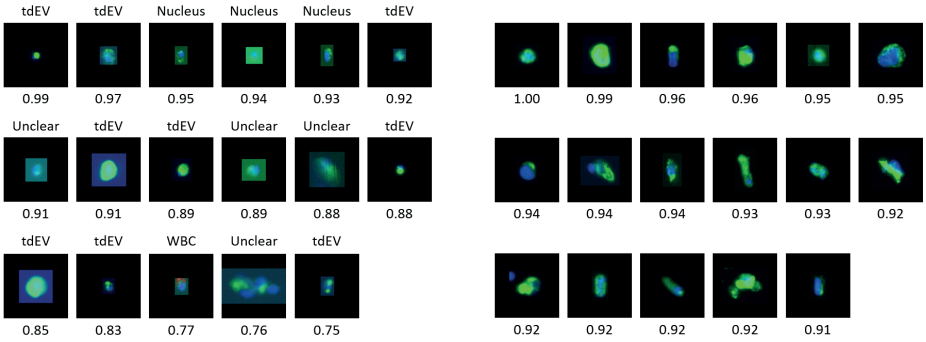
In our comparison between the manual and the automatic counts and their relation to patient survival, we only classified the CK+/DAPI+ CTC candidates presented by the CellSearch® system. Yet, to use the full power of our multi-class classification approach, we need to present all events found in a cartridge to the network, regardless of their marker expression, size and morphology. In that way, we obtain not only a CTC count per sample but also counts of tdEVS, WBCs, bare nuclei, a novel population of events identified by our autoencoding network which we classified as extracellular vesicles derived from white blood cells (CD45EVs) and all other objects present in the images. This provides us with valuable information that might help to give a better prediction for survival outcome and success of a therapy. It can also alleviate the problem that CTCs are very rare and difficult to detect and missing one single CTC can lead to the misclassification of a patient as having a relatively good prognosis. In [37] Nanou et al. showed that tdEV counts can also be used as a prognostic marker equivalent to CTC counts. The reason for this might be that tdEVs are either released by the primary or the secondary tumor in a similar way as CTCs or originate from apoptotic CTCs resulting in a correlation of both counts. But for other cell classes like bare nuclei, it is so far not investigated whether their count could add to the predictive value of a test, although we have shown that they are also present in significantly higher numbers in patient samples compared to healthy donors [10]. Alternatively, it could be also the

case that not all CTCs are equaly predictive for survival and that only cells with a very high probability of being a CTC should be incorporated in the analysis. Using the autoencoded classification network and visualizing the latent space might also reveal subclasses of CTCs or other cells that are more predictive than other subclassses. A clear example of subclasses identified by the autoencoding network can be observed in the tdEVs illustrated in Figure 6.5 (d), where the arrows indicate the presence of eight distinct populations of tdEVs. In order to evaluate the clinical relevance of each subclass, their association to the clinical outcome of patients and their downstream characterization in terms of their molecular cargo need to be investigated.

To detect all events in full scans for further classification, the ACCEPT toolbox can be used. It includes a multi-scale segmentation method [61] to robustly detect all signals in each of the fluorescent channels. Currently, to classify all found events, we need to export them from the ACCEPT toolbox programmed in MATLAB® (Mathworks, Natick, MA), save them on the disc and reload them for classification by the network, using Keras. For samples with up-to 20,000 events this is feasible, yet not very efficient, but for samples with more detected events this approach is not feasible. Therefore, to efficiently use our trained DL network, we plan to integrate it in the toolbox. Once this is done, the ACCEPT toolbox can be used to iteratively enlarge the training database by using the found class members once they are manually reviewed. Retraining the network on an iteratively growing database will improve the robustness and enlarge the variety of cells the network has seen. In this way, the given scores will gain reliability and manual revision becomes redundant resulting in a fully automated approach for CTC detection that standardizes the definition of a CTC. As a consequence of first segmenting all objects and then classifying them with the network, the classification quality will be always biased by the quality of the preceding segmentation. If the segmentation fails to detect a cell, it will not be presented to the network and therefore not counted in the respective class. In [10] we have shown that this is often a problem in case of either very full samples with up-to 500,000 cells, or in case of an illumination bias in the images. An alternative would be to use a “sliding window” approach where a window of the network’s input size slides over the whole frame and classifies the part of the image currently covered by the window. A disadvantage of that approach is that several cells might be presented to the network at once and assigning a single class is not possible. Also, a reduced input size might not cover large events, especially artifacts or clusters found in the fluorescent images. Moreover, for samples with lower overall cell counts, most of the image is just background whose classification will only increase the run time although no classification would be needed if a reliable preselection would be made. A promising approach especially for samples with a low image quality is to use another convolutional network for the segmentation [44, 10] and ultimately couple the segmentation and the classification to a semantic segmentation network [33].

CTC hold the promise to represent a real-time liquid biopsy provided they are present and their content can be characterized. Adding the information of the genetic and proteomic make-up of each CTC would truly enable to choose the optimal therapy for the patient at any given time point during the course of the disease. The complexity of this information surely will benefit by the use of deep learning methods. A first step could simply be to provide

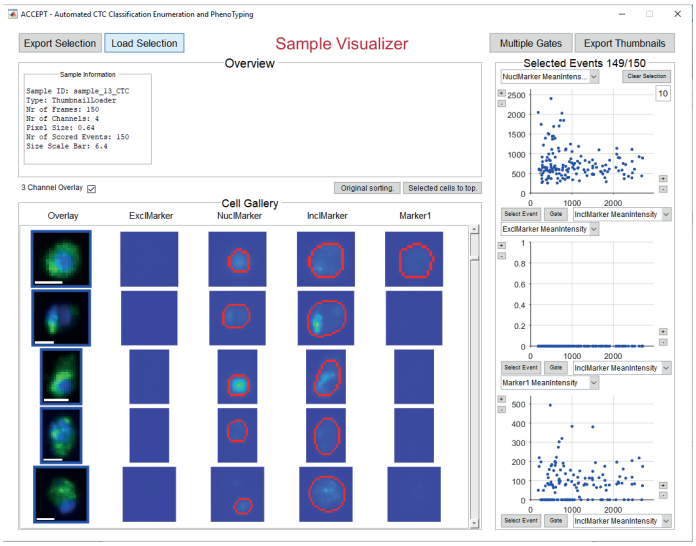
a likelihood for an identified CTC to yield high quality DNA that can be analyzed. A second step could be to identify targets for a treatment and a third step could be to perform longitudinal measurements and identify changes which can be a result of treatment resistance. Ultimately the algorithms should tell us when to start which treatment and when to switch to an alternative therapy for prolonged survival of the patient.



(a) Cells labeled as a CTC but rejected as such by the network.

(b) Cells not labeled as a CTC but classified as such by the network.

Supplementary Figure 6.1: Overview of cells that were mistakenly scored as a CTC or rejected as such by the 6-class network. Panel (a) shows all cells that were labeled as a CTC but were assigned by the network to a different class (reported above the thumbnail image) with a probability of over 75%. Panel (b) shows all cells that were not labeled as a CTC (although this might be questionable) but classified as a CTC by the network with a probability of over 90%. For each cell, the assigned class probability is reported below the thumbnail image.



Supplementary Figure 6.2: Screenshot of the ACCEPT visualization tool that was used for the generation of our ground truth set. By default, all candidates were selected to belong to the predefined cell class. By going through the cell gallery and clicking on the overlay image on the left, candidates can be rejected to belong to the cell class.

Bibliography

[1] J. C. Caicedo, S. Cooper, F. Heigwer, S. Warchal, P. Qiu, C. Molnar, A. S. Vasilevich, J. D. Barry, H. S. Bansal, O. Kraus, M. Wawer, L. Paavolainen, M. D. Herrmann, M. Rohban, J. Hung, H. Hennig, J. Concannon, I. Smith, P. A. Clemons, S. Singh, P. Rees, P. Horvath, R. G. Linington, and A. E. Carpenter. Data-Analysis Strategies for Image-Based Cell Profiling. *Nature Methods*, 14(9):849–863, aug 2017.

[2] C. L. Chen, A. Mahjoubfar, L. C. Tai, I. K. Blaby, A. Huang, K. R. Niazi, and B. Jalali. Deep Learning in Label-free Cell Classification. *Scientific Reports*, 6:21471, 2016.

[3] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. DeCaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter, and C. S. Greene. Opportunities and Obstacles for Deep Learning in Biology and Medicine. *Journal of The Royal Society Interface*, 15(141):20170387, apr 2018.

[4] F. Chollet. Keras, 2015.

[5] S. J. Cohen, C. J. Punt, N. Iannotti, B. H. Saidman, K. D. Sabbath, N. Y. Gabrail, J. Picus, M. Morse, E. Mitchell, M. C. Miller, G. V. Doyle, H. Tissing, L. W. Terstappen, and N. J. Meropol. Relationship of Circulating Tumor Cells to Tumor Response, Progression-Free Survival, and Overall Survival in Patients with Metastatic Colorectal Cancer. *Journal of Clinical Oncology*, 26:3213–3221, 2008.

[6] F. A. Coumans, S. T. Ligthart, and L. W. Terstappen. Interpretation of Changes in Circulating Tumor Cell Counts. *Translational Oncology*, 5(6):486–491, 2012.

[7] M. Crespo, G. van Dalum, R. Ferraldeschi, Z. Zafeiriou, S. Sideris, D. Lorente, D. Bianchini, D. Nava-Rodrigues, R. Riisnaes, S. Miranda, I. Figueiredo, P. Flohr, K. Nowakowska, J. S. de Bono, L. W. Terstappen, and G. Attard. Androgen Receptor Expression in Circulating Tumour Cells from Castration-Resistant Prostate Cancer Patients Treated with Novel Endocrine Agents. *British Journal of Cancer*, 112(7):1166, 2015.

[8] M. Cristofanilli, G. T. Budd, M. J. Ellis, A. Stopeck, J. Matera, M. C. Miller, J. M. Reuben, G. V. Doyle, W. J. Allard, L. W. Terstappen, and D. F. Hayes. Circulating Tumor Cells, Disease Progression, and Survival in Metastatic Breast Cancer. *New England Journal of Medicine*, 351(8):781–791, 2004.

[9] J. S. de Bono, H. I. Scher, R. B. Montgomery, C. Parker, M. C. Miller, H. Tissing, G. V. Doyle, L. W. Terstappen, K. J. Pienta, and D. Raghavan. Circulating Tumor Cells Predict Survival Benefit From Treatment in Metastatic Castration-Resistant Prostate Cancer. *Clinical Cancer Research*, 14(19):6302–6309, 2008.

[10] S. de Wit. *Circulating Tumor Cells and Beyond*. PhD thesis, University of Twente, Enschede, The Netherlands, jun 2018.

[11] S. de Wit, G. van Dalum, A. T. Lenferink, A. G. Tibbe, T. J. N. Hiltermann, H. J. Groen, C. J. van Rijn, and L. W. Terstappen. The Detection of EpCAM+ and EpCAM- Circulating Tumor Cells. *Scientific Reports*, 5:12270, 2015.

[12] O. Dürr and B. Sick. Single-Cell Phenotype Classification Using Deep Convolutional Neural Networks. *Journal of Biomolecular Screening*, 21(9):998–1003, oct 2016.

[13] D. Erhan, A. Courville, and P. Vincent. Why Does Unsupervised Pre-training Help Deep Learning ? *Journal of Machine Learning Research*, 11(Feb):625–660, 2010.

[14] P. Eulenberg, N. Köhler, T. Blasi, A. Filby, A. E. Carpenter, P. Rees, F. J. Theis, and F. A. Wolf. Reconstructing Cell Cycle and Disease Progression Using Deep Learning. *Nature Communications*, 8(1):463, dec 2017.

[15] T. Fehm, V. Müller, B. Aktas, W. J. Janni, A. Schneeweiss, E. Stickeler, C. Lattrich, C. R. Löhberg, E. Solomayer, B. Rack, S. Riethdorf, C. Klein, C. Schindlbeck, K. Brocker, S. Kasimir-Bauer, D. Wallwiener, and K. Pantel. HER2 Status of Circulating Tumor Cells in Patients with Metastatic Breast Cancer: A Prospective, Multicenter Trial. *Breast Cancer Research and Treatment*, 124(2):403–412, 2010.

[16] K. Fizazi, H. I. Scher, A. Molina, C. J. Logothetis, K. N. Chi, R. J. Jones, J. N. Staffurth, S. North, N. J. Vogelzang, F. Saad, P. Mainwaring, S. Harland, O. B. Goodman, C. N. Sternberg, J. H. Li, T. Kheoh, C. M. Haqq, and J. S. de Bono. Abiraterone Acetate for Treatment of Metastatic Castration-Resistant Prostate Cancer: Final Overall Survival Analysis of the COU-AA-301 Randomised, Double-Blind, Placebo-Controlled Phase 3 Study. *The Lancet Oncology*, 13(10):983–992, 2012.

[17] B. Franken, M. R. de Groot, W. J. Mastboom, I. Vermes, J. van der Palen, A. G. Tibbe, and L. W. Terstappen. Circulating Tumor Cells, Disease Recurrence and Survival in Newly Diagnosed Breast Cancer. *Breast Cancer Research*, 14(5):R133, 2012.

[18] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li. Deep Reconstruction-Classification Networks for Unsupervised Domain Adaptation. In *European Conference on Computer Vision*, pages 597–613, 2016.

[19] H. Greenspan, B. van Ginneken, and R. M. Summers. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging*, 35(5):1153–1159, may 2016.

[20] D. F. Hayes, M. Cristofanilli, G. T. Budd, M. J. Ellis, A. Stopeck, M. C. Miller, J. Matera, W. J. Allard, G. V. Doyle, and L. W. Terstappen. Circulating Tumor Cells at each Follow-Up Time Point During Therapy of Metastatic Breast Cancer Patients Predict Progression-Free and Overall Survival. *Clinical Cancer Research*, 12(14):4218–4224, 2006.

[21] G. E. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury. Deep Neural Networks for Acoustic Modeling in Speech Recognition. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[22] M. Ignatiadis, S. Riethdorf, F.-C. Bidard, I. Vaucher, M. Khazour, F. Rothé, J. Metallo, G. Rouas, R. E. Payne, R. Coombes, I. Teufel, U. Andergassen, S. Apostolaki, E. Politaki, D. Mavroudis, S. Bessi, M. Pestrin, A. Di Leo, M. Campion, M. Reinholz, E. Perez, M. Piccart,

E. Borgen, B. Naume, J. Jimenez, C. Aura, L. Zorzino, M. Cassatella, M. Sandri, B. Mostert, S. Sleijfer, J. Kraan, W. J. Janni, T. Fehm, B. Rack, L. W. Terstappen, M. I. Repollet, J.-Y. Pierga, M. C. Miller, C. Sotiriou, S. Michiels, and K. Pantel. International Study on Inter-Reader Variability for Circulating Tumor Cells in Breast Cancer. *Breast Cancer Research*, 16(2):R43, 2014.

[23] A. Kapil, A. Meier, A. Zuraw, K. Steele, M. Rebelatto, G. Schmidt, and N. Brieu. Deep Semi Supervised Generative Learning for Automated PD-L1 Tumor Cell Scoring on NSCLC Tissue Needle Biopsies. *arXiv:1806.11036*, jun 2018.

[24] J. Ko, S. N. Baldassano, P.-L. Loh, K. Kording, B. Litt, and D. Issadore. Machine Learning to Detect Signatures of Disease in Liquid Biopsies – a User's Guide. *Lab on a Chip*, 18(3):395–405, 2018.

[25] J. Kraan, S. Sleijfer, M. H. Strijbos, M. Ignatiadis, D. Peeters, J.-Y. Pierga, F. Farace, S. Riethdorf, T. Fehm, L. Zorzino, A. G. Tibbe, M. Maestro, R. Gisbert-Criado, G. Denton, J. S. de Bono, C. Dive, J. A. Foekens, and J. W. Gratama. External Quality Assurance of Circulating Tumor Cell Enumeration Using the CellSearch® System: A Feasibility Study. *Cytometry Part B*, 80B(2):112–118, 2011.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances In Neural Information Processing Systems*, pages 1097–1105, 2012.

[27] P. Lambin, R. T. Leijenaar, T. M. Deist, J. Peerlings, E. E. de Jong, J. van Timmeren, S. Sanduleanu, R. T. Larue, A. J. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. M. Mottaghy, J. E. Wildberger, and S. Walsh. Radiomics: the Bridge Between Medical Imaging and Personalized Medicine. *Nature Reviews Clinical Oncology*, 14(12):749, oct 2017.

[28] T. B. Lannin, F. I. Thege, and B. J. Kirby. Comparison and Optimization of Machine Learning Methods for Automated Classification of Circulating Tumor Cells. *Cytometry Part A*, 89(10):922–931, oct 2016.

[29] Y. A. LeCun, Y. Bengio, and G. E. Hinton. Deep Learning. *Nature*, 521(7553):436, 2015.

[30] G. Litjens, T. Kooi, B. Ehteshami Bejnordi, A. Arindra Adiyoso Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sanchez. A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, 42:60–88, 2017.

[31] G. Litjens, C. I. Sánchez, N. Timofeeva, M. Hermsen, I. D. Nagtegaal, I. Kovacs, C. Hulsbergen - van de Kaa, P. Bult, B. van Ginneken, and J. A. van der Laak. Deep Learning as a Tool for Increased Accuracy and Efficiency of Histopathological Diagnosis. *Scientific Reports*, 6(1):26286, sep 2016.

[32] J. Liu, B. Xu, L. Shen, J. Garibaldi, and G. Qiu. HEp-2 Cell Classification Based on a Deep Autoencoding-Classification Convolutional Neural Network. In *IEEE International Symposium on Biomedical Imaging*, pages 1019–1023. IEEE, 2017.

[33] J. Long, E. Shelhamer, and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[34] Y. Mao, Z. Yin, and J. M. Schober. Iteratively Training Classifiers for Circulating Tumor Cell Detection. In *IEEE International Symposium on Biomedical Imaging*, pages 190–194. IEEE, apr 2015.

[35] Y. Mao, Z. Yin, and J. M. Schober. A Deep Convolutional Neural Network Trained on Representative Samples for Circulating Tumor Cell Detection. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–6. IEEE, mar 2016.

[36] G. Montavon, W. Samek, and K.-R. Müller. Methods for Interpreting and Understanding Deep Neural Networks. *Digital Signal Processing*, 73:1–15, feb 2018.

[37] A. Nanou, F. A. Coumans, G. van Dalum, L. L. Zeune, D. Dolling, W. Onstenk, M. Crespo, M. Sousa Fontes, P. Rescigno, G. Fowler, P. Flohr, C. Brune, S. Sleijfer, J. S. de Bono, and L. W. Terstappen. Circulating Tumor Cells, Tumor-Derived Extracellular Vesicles and Plasma Cytokeratins in Castration-Resistant Prostate Cancer Patients. *Oncotarget*, 9(27):19283–19293, 2018.

[38] J.-Y. Pierga, T. Petit, T. Delozier, J. M. Ferrero, M. Campone, J. Gligorov, F. Lerebours, H. Roché, T. Bachelot, E. Charafe-Jauffret, M. Pavlyuk, S. Kraemer, F.-C. Bidard, and P. Viens. Neoadjuvant Bevacizumab, Trastuzumab, and Chemotherapy for Primary Inflammatory HER2-Positive Breast Cancer (BEVERLY-2): An Open-Label, Single-Arm Phase 2 Study. *The Lancet Oncology*, 13(4):375–384, 2012.

[39] J.-Y. Pierga, T. Petit, C. Lévy, J. M. Ferrero, M. Campone, J. Gligorov, F. Lerebours, H. Roché, T. Bachelot, E. Charafe-Jauffret, J. Bonneterre, J. Hernandez, F.-C. Bidard, and P. Viens. Pathological Response and Circulating Tumor Cell Count Identifies Treated HER2+ Inflammatory Breast Cancer Patients with Excellent Prognosis: BEVERLY-2 Survival Data. *Clinical Cancer Research*, 21(6):1298–1304, 2015.

[40] C. B. Raub and G. Nehmetallah. Holography, Machine Learning, and Cancer Cells. *Cytometry Part A*, 91(8):754–756, aug 2017.

[41] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang. Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics*, 21(1):4–21, jan 2017.

[42] A. Regev, S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, H. Clevers, B. Deplancke, I. Dunham, J. Eberwine, R. Eils, W. Enard, A. Farmer, L. Fugger, B. Göttgens, N. Hacohen, M. Haniffa, M. Hemberg, S. Kim, P. Klenerman, A. Kriegstein, E. Lein, S. Linnarsson, E. Lundberg, J. Lundeberg, P. Majumder, J. C. Marioni, M. Merad, M. Mhlanga, M. Nawijn, M. Netea, G. Nolan, D. Pe'er, A. Phillipakis, C. P. Ponting, S. Quake, W. Reik, O. Rozenblatt-Rosen, J. Sanes, R. Satija, T. N. Schumacher, A. Shalek, E. Shapiro, P. Sharma, J. W. Shin, O. Stegle, M. Stratton, M. J. Stubbington, F. J. Theis, M. Uhlen, A. van Oudenaarden, A. Wagner, F. Watt, J. Weissman, B. Wold, R. Xavier, and N. Yosef. The Human Cell Atlas: From Vision to Reality. *eLife*, 550(7677):451, dec 2017.

[43] S. Robertson, H. Azizpour, K. Smith, and J. Hartman. Digital Image Analysis in Breast Pathology - from Image Processing Techniques to Artificial Intelligence. *Translational Research*, 194:19–35, 2018.

[44] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[45] M. Saha and C. Chakraborty. HER2Net: A Deep Framework for Semantic Segmentation and Classification of Cell Membranes and Nuclei in Breast Cancer Evaluation. *IEEE Transactions on Image Processing*, 27(5):2189–2200, may 2018.

[46] C.-M. Svensson, R. Hübler, and M. T. Figge. Automated Classification of Circulating Tumor Cells and the Impact of Interobsever Variability on Classifier Training and Performance. *Journal of Immunology Research*, 2015:573165, 2015.

[47] C.-M. Svensson, S. Krusekopf, J. Lücke, and M. T. Figge. Automated Detection of Circulating Tumor Cells with Naive Bayesian Classifiers. *Cytometry Part A*, 85(6):501–511, jun 2014.

[48] J. Tol, M. Koopman, M. C. Miller, A. G. Tibbe, A. Cats, G.-J. M. Creemers, A. H. Vos, I. D. Nagtegaal, L. W. Terstappen, and C. J. Punt. Circulating Tumour Cells Early Predict Progression-Free and Overall Survival in Advanced Colorectal Cancer Patients Treated with Chemotherapy and Targeted Agents. *Annals of Oncology*, 21(5):1006–1012, 2009.

[49] L. J. van der Maaten and G. E. Hinton. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008.

[50] M. E. Vandenberghe, M. L. Scott, P. W. Scorer, M. Söderberg, D. Balcerzak, and C. Barker. Relevance of Deep Learning to Facilitate the Diagnosis of HER2 Status in Breast Cancer. *Scientific Reports*, 7:45938, dec 2017.

[51] R. Vidal, J. Bruna, R. Giryes, and S. Soatto. Mathematics of Deep Learning. *arXiv:1712.04741*, 2017.

[52] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol. Extracting and Composing Robust Features with Denoising Autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103, 2008.

[53] G. Wang. A Perspective on Deep Imaging. *IEEE Access*, 4:8914–8924, 2016.

[54] F. A. Wolf, P. Angerer, and F. J. Theis. SCANPY: Large-Scale Single-Cell Gene Expression Data Analysis. *Genome Biology*, 19(1):15, dec 2018.

[55] S. Yang, B. Fang, W. Tang, X. Wu, J. Qian, and W. Yang. Faster R-CNN Based Microscopic Cell Detection. In *2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC)*, pages 345–350. IEEE, dec 2017.

[56] F. Yellin, B. D. Haeffele, and S. Roth. Multi-Cell Detection and Classification Using a Generative Convolutional Model. In *CVPR 2018*, pages 8953–8961, 2018.

[57] M. D. Zeiler. ADADELTA: An Adaptive Learning Rate Method. *arXiv:1212.5701*, 2012.

[58] M. D. Zeiler and R. Fergus. Visualizing and Understanding Convolutional Networks. In *European Conference on Computer Vision*, pages 818–833. Springer, 2014.

[59] L. L. Zeune, S. de Wit, A. S. Berghuis, M. J. IJzerman, L. W. Terstappen, and C. Brune. How to Agree on a CTC. *Cytometry Part A*, 2018.

[60] L. L. Zeune, G. van Dalum, C. Decraene, C. Proudhon, T. Fehm, H. Neubauer, B. Rack, M. Alunni-Fabbroni, L. W. Terstappen, S. A. van Gils, and C. Brune. Quantifying HER-2 Expression on Circulating Tumor Cells by ACCEPT. *PLOS ONE*, 12(10):e0186562, oct 2017.

[61] L. L. Zeune, G. van Dalum, L. W. Terstappen, S. A. van Gils, and C. Brune. Multiscale Segmentation via Bregman Distances and Nonlinear Spectral Analysis. *SIAM Journal on Imaging Sciences*, 10(1):111–146, 2017.

[62] J. Zhao, M. Zhang, Z. Zhou, J. Chu, and F. Cao. Automatic Detection and Classification of Leukocytes Using Convolutional Neural Networks. *Medical & Biological Engineering & Computing*, 55(8):1287–1301, aug 2017.

Conclusion and Outlook

The research presented in this thesis is focused on the development of image analysis and machine learning tools for automated and reliable detection and classification of Circulating Tumor Cells (CTCs) in fluorescence microscopy images. CTCs are found in the blood of cancer patients and are a promising biomarker in precision medicine. Yet, to fully exploit the potential of CTCs, automated image analysis and classification tools are needed. We addressed this need with the development of the software tool "ACCEPT". The toolbox is based on the work presented in the six chapters of this thesis which can be subdivided into three different categories:

1. Image analysis models for robust cell segmentation (Chapter 1 & 2)
2. Biomedical applications and new findings based on quantifiable image features (Chapter 3 & 4)
3. Analysis of manual CTC scoring and its automation by machine learning methods (Chapter 5 & 6).

In Chapter 1, we introduced a novel segmentation approach, which couples an active-contour segmentation model [3] with a nonlinear spectral analysis [8, 2]. The resulting segmentation model is very robust towards noise and parameter changes. This means that no extensive parameter tuning is needed to obtain segmentation results of reliable quality [17]. Moreover, the spectral analysis provides useful insight into the spectral composition of the image in terms of size and intensity scales. Our segmentation method is well-suited for the automated segmentation of cells in fluorescent images and forms the basis of our toolbox for CTC analysis. In Chapter 2, we adapted the model to make it an intensity-independent, size-based segmentation model, which outperforms the previous model in segmenting very dim intensity signals [18].

In Chapter 3, we evaluated the assessment of expression of treatment targets on CTCs with and without the ACCEPT toolbox. This was tested on a cohort of 132 metastatic breast cancer patients from which blood samples were processed and stained with HER-2 as an additional marker [7, 13]. Using ACCEPT, we analyzed all manually scored CTCs and showed that the extraction and analysis of quantitative image features in combination with a reliable visualization of the fluorescent signal improves the agreement and reproducibility of scoring

CTCs positive or negative for additional fluorescent markers [16]. In Chapter 4, we used the semi-automated classification of ACCEPT to detect multiple cell populations. Our analysis revealed that, apart from CTCs, also other cell populations are overly represented in patient blood samples compared to healthy donors. It remains for future research to investigate whether these cells can be used for a better prediction of prognosis. Moreover, we showed how deep learning based segmentation methods can overcome difficulties in reliable segmentation of cells in case of strong illumination biases or cartridges that are overcrowded with cells [4].

In Chapter 5, we evaluated the current consensus in scoring objects as CTCs. We created a set of 100 cells and asked 15 trained reviewers whether or not each cell is a CTC, and evaluated the consensus. Moreover, we compared the results with the scores of a panel of four experts in the field and with the results of a Deep Learning (DL) network trained to classify CTCs. Computing the agreement for each two of the three methods showed in all cases moderate to substantial agreement, but was the highest for the 15 reviewers with the DL network [15]. This strongly motivated the use of DL methods for CTC analysis presented in Chapter 6. Here, we presented a novel, fully-automated DL approach for classification and analysis of CTCs and other cell populations found in fluorescence images, obtained after immunomagnetic enrichment of CTCs, from blood of cancer patients. Apart from the very accurate and robust identification of CTCs, the method allows for the exploration of new cell populations and cell subclasses based on advanced visualization techniques. Moreover, on a set of 100 metastatic breast cancer patients, our method excelled in predicting the overall survival of cancer patients compared to state-of-the-art manual counts.

Besides the work presented in this thesis, a major part of this PhD project was dedicated to the development of the open-source toolbox for CTC analysis. The goal of this toolbox is to provide the research field of liquid biopsies with a user-friendly, free-to-use but advanced tool for the automated analysis of fluorescence images of CTCs. The “ACCEPT - Automated CTC Classification, Enumeration and PhenoTyping” toolbox can be downloaded from *https://github.com/LeonieZ/ACCEPT*. The toolbox can load and analyze images from multiple fluorescence microscope systems commonly used in the field. ACCEPT can either be used to analyze manually pre-scored CTCs for additional marker expression for treatment decisions [16] or alternatively to detect and analyze all cells and objects present in the fluorescent images of the blood. For each cell, multiple quantitative features based on size and morphology of the cell are extracted and can be used for semi-automated cell classification or manual analysis of all cells. By using the semi-automated detection and classification of all cells found in the images, researchers can analyze not only the amount of CTCs found in the blood of cancer patients, but also the amount of representatives for other cell classes, which is provably also related to the survival chance of patients [12, 6].

Future Perspective This thesis, and especially the last chapter on Deep Learning of CTCs, opened the way for multiple new short-term and long-term projects. The following paragraph starts with the short-terms projects and moves towards our vision for the ACCEPT project.

In Chapter 6 we observed, that the quality of the automated classification of CTCs varies per cancer type and is correlated to the number of examples in the ground-truth set. This motivates the extension of the ground-truth set by cells from under-represented cancer types. Yet, manually scoring CTCs is tedious and time-intensive especially for cancer types with very few CTCs, like non-small cell lung cancer [5]. Therefore, to efficiently enlarge the ground-truth set also with examples from different microscope systems, we plan to couple the ACCEPT toolbox with the DL classification and an export function so that users can easily export the images once they manually reviewed the DL scores and have corrected them if necessary. In that way the classification will be iteratively improved and become more and more robust towards different machines and cancer types.

The analysis of the latent space of the DL network in Chapter 6 revealed the presence of multiple subclasses of the cells. Yet, so far, it is unclear whether these newly identified subclasses have any biological or clinical implications. It remains for further analysis to investigate whether the subclasses have different impact on the progression of cancer or the survival chances of a patient. A first step would be to extract the number of cell counts for each subclass and perform a Kaplan-Meier analysis to see whether their number is related to the outcome of patients or not.

In Chapter 1 and 2 we have shown that nonlinear spectral analysis [8, 2] can be used to automatically extract information about size and intensity scales of cells and their composition in the image. Yet, Chapter 4 demonstrated that the segmentation quality of these methods decreases with lower image and sample quality. In these cases, Deep Learning based segmentation [14] outperformed the classical methods. The architecture of the network we used is related to the architecture of the network we used in Chapter 6. As a first step, one could combine both networks so that the segmentation and the classification of cells is addressed by a single model for semantic segmentation [11]. The analysis of the latent space in Chapter 6 has shown us that scale information in terms of size and intensity of the fluorescent marker of a cell are encoded in the latent space. This is reminiscent of the encoded scale information of the spectral response function of Chapter 1. Indeed, it can be shown that autoencoders used in Chapter 6 are related to nonlinear diffusion methods used in Chapter 1 [9]. This motivates the development of a framework which couples spectral analysis and scale decomposition with autoencoding neural networks to combine both our works.

Our current definition of a tumor cell is solely based on the fluorescent images of the cells. Yet, to obtain the “true” answer whether a cell really is a CTC, it requires the molecular information whether genetic aberrations are present, confirming that the cell is a CTC. If the molecular information and the corresponding fluorescent images are available for multiple cells, we could investigate whether certain image features are linked to the answer whether a cell really is a CTC or not. In that way, a further discrimination of the cells in the ground-truth set for Deep Learning could be achieved. Besides, the genetic information of a CTC after whole genome amplification can be used to find specific mutations that can be utilized to guide the therapy of cancer patients. Yet, the procedure of whole genome amplification is very tedious and expensive and many cells are lost or turn out to be not suitable for DNA amplification. Therefore, we are interested in finding out whether certain image features

found by the network or the corresponding score can be related to the ability of a cell to be suitable for whole genome amplification. If this is not possible right now, maybe the network could be retrained to include that information, once enough ground-truth data is available.

Another research direction in line with hybrid imaging techniques on organ-level like PET-CT [1] or PET-MRI [10] is the coupling of many imaging sources of one cell. Currently, the ACCEPT toolbox is specialized for the analysis of fluorescence microscopy images. Yet, imaging techniques like Scanning Electron Microscopy (SEM) or Raman spectroscopy provide extra information not captured by fluorescence microscopy. Extending the toolbox to provide image analysis and machine learning tools for additional imaging techniques has the potential to further explore different cell classes and their clinical implications based on expanded insight we gain into single cells.

The ultimate goal for the ACCEPT toolbox is to bring it into the clinics and use it for prognosis and treatment decisions of cancer patients. In that way, the field can profit from a unified definition of a CTC based on advanced imaging and machine learning tools. The user bias in scoring CTCs as we have seen in Chapter 5 would be eliminated and treatment decisions would be based on a subjective CTC count. If the detection and classification of all cells is fully automated, also the training in scoring CTCs and the time spent with manual scoring can be eliminated. Using the ACCEPT toolbox for automated cell counts also allows to move away from static data-sets but monitor the cell counts over time. In that way, the ACCEPT toolbox can be used to analyze when a “true” change in cell counts has happened and can guide the decision whether a therapy is still helpful or when a therapy should be changed. Apart from economical benefits, an automated early answer to this question has the potential to improve the quality of life for cancer patients.

Bibliography

[1] T. Beyer, D. W. Townsend, T. Brun, P. E. Kinahan, M. Charron, R. Roddy, J. Jerin, J. Young, L. Byars, and R. Nutt. A Combined PET/CT Scanner for Clinical Oncology. *Journal of Nuclear Medicine*, 41(8):1369–1379, 2000.

[2] M. Burger, L. Eckardt, G. Gilboa, and M. Moeller. Spectral Representations of One-Homogeneous Functionals. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 16–27. Springer, 2015.

[3] T. F. Chan and L. A. Vese. Active Contours Without Edges. *IEEE Transactions on Image Processing*, 10(2):266–277, 2001.

[4] S. de Wit. *Circulating Tumor Cells and Beyond*. PhD thesis, University of Twente, Enschede, The Netherlands, jun 2018.

[5] S. de Wit, G. van Dalum, and L. W. Terstappen. Detection of Dirculating Tumor Cells. *Scientifica*, 2014:819362, 2014.

[6] S. de Wit, L. L. Zeune, T. J. N. Hiltermann, H. J. Groen, G. van Dalum, and L. W. Terstappen. Classification of Cells in CTC Enriched Samples by Advanced Image Analysis. *Cancers*, 10(10):E377, 2018.

[7] T. Fehm, V. Müller, B. Aktas, W. J. Janni, A. Schneeweiss, E. Stickeler, C. Lattrich, C. R. Löhberg, E. Solomayer, B. Rack, S. Riethdorf, C. Klein, C. Schindlbeck, K. Brocker, S. Kasimir-Bauer, D. Wallwiener, and K. Pantel. HER2 Status of Circulating Tumor Cells in Patients with Metastatic Breast Cancer: A Prospective, Multicenter Trial. *Breast Cancer Research and Treatment*, 124(2):403–412, 2010.

[8] G. Gilboa. A Total Variation Spectral Framework for Scale and Texture Analysis. *SIAM Journal on Imaging Sciences*, 7(4):1937–1961, jan 2014.

[9] N. Huttinga. *Insights into Deep Learning Methods with Application to Cancer Imaging*. Master thesis, University of Twente, 2017.

[10] M. S. Judenhofer, H. F. Wehrl, D. F. Newport, C. Catana, S. B. Siegel, M. Becker, A. Thielscher, M. Kneilling, M. P. Lichy, M. Eichner, K. Klingel, G. Reischl, S. Widmaier, M. Röcken, R. E. Nutt, H. J. Machulla, K. Uludag, S. R. Cherry, C. D. Claussen, and B. J. Pichler. Simultaneous PET-MRI: A New Approach for Functional and Morphological Imaging. *Nature Medicine*, 14(4):459, 2008.

[11] J. Long, E. Shelhamer, and T. Darrell. Fully Convolutional Networks for Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440, 2015.

[12] A. Nanou, F. A. Coumans, G. van Dalum, L. L. Zeune, D. Dolling, W. Onstenk, M. Crespo, M. Sousa Fontes, P. Rescigno, G. Fowler, P. Flohr, C. Brune, S. Sleijfer, J. S. de Bono, and L. W. Terstappen. Circulating Tumor Cells, Tumor-Derived Extracellular Vesicles and Plasma Cytokeratins in Castration-Resistant Prostate Cancer Patients. *Oncotarget*, 9(27):19283–19293, 2018.

[13] J.-Y. Pierga, T. Petit, T. Delozier, J. M. Ferrero, M. Campone, J. Gligorov, F. Lerebours, H. Roché, T. Bachelot, E. Charafe-Jauffret, M. Pavlyuk, S. Kraemer, F.-C. Bidard, and P. Viens. Neoadjuvant Bevacizumab, Trastuzumab, and Chemotherapy for Primary Inflammatory HER2-Positive Breast Cancer (BEVERLY-2): An Open-Label, Single-Arm Phase 2 Study. *The Lancet Oncology*, 13(4):375–384, 2012.

[14] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

[15] L. L. Zeune, S. de Wit, A. S. Berghuis, M. J. IJzerman, L. W. Terstappen, and C. Brune. How to Agree on a CTC. *Cytometry Part A*, 2018.

[16] L. L. Zeune, G. van Dalum, C. Decraene, C. Proudhon, T. Fehm, H. Neubauer, B. Rack, M. Alunni-Fabbroni, L. W. Terstappen, S. A. van Gils, and C. Brune. Quantifying HER-2 Expression on Circulating Tumor Cells by ACCEPT. *PLOS ONE*, 12(10):e0186562, oct 2017.

[17] L. L. Zeune, G. van Dalum, L. W. Terstappen, S. A. van Gils, and C. Brune. Multiscale Segmentation via Bregman Distances and Nonlinear Spectral Analysis. *SIAM Journal on Imaging Sciences*, 10(1):111–146, 2017.

[18] L. L. Zeune, S. A. van Gils, L. W. Terstappen, and C. Brune. Combining Contrast Invariant L1 Data Fidelities with Nonlinear Spectral Image Decomposition. In *Lecture Notes in Computer Science*, pages 80–93, 2017.

C

Summary

Cancer is one of the leading causes of death worldwide. It starts with the formation of a primary tumor and can spread throughout the body, which ultimately causes most cancer related deaths. Tumor cells that break away from the tumor and invade the bloodstream are called Circulating Tumor Cells (CTCs). Extravasation of these CTCs can give rise to new tumors at distant sites and this process is generally referred to as the formation of metastases. The CTC load is strongly correlated with the time of survival of cancer patients and the number of CTCs can be used to monitor cancer therapy. Yet, CTCs are very rare and to accurately detect, characterize and count them in images of fluorescently labeled cells is very challenging and often performed manually. This process is prone to errors and user biases. Therefore, this thesis aims to develop image analysis and machine learning models to automatically and accurately detect and classify CTCs in fluorescent images and prove the added benefit of automated image analysis methods for patients' benefit.

CTCs and their fluorescent signals strongly vary in size, shape and signal intensity. We addressed the presence of multiple size and intensity scales in the fluorescent images by the development of a multiscale segmentation model. The model is almost parameter-independent and extracts the scale information based on a spectral decomposition of the input signal using nonlinear diffusion models. This results in an accurate segmentation of the cells combined with a clustering of the cells based on their size and intensity. We further adapted the model to a purely size-based clustering of the cells which excels in segmenting also very dim fluorescent signals.

Based on the segmentation of all cells present in the images, we can extract quantifiable features per cell, such as their size or mean fluorescent intensity. Providing these extracted features to researchers who score the fluorescent expression of CTCs as positive or negative, could greatly reduce the user bias and unify the results used for patient treatment. Moreover, we used the extracted features to identify new cell populations which were also over-expressed by cancer patients compared to healthy subjects. Identifying and understanding these cell populations has the potential to better understand the variation among patients in disease progression and treatment response.

We further analyzed the consensus of multiple reviewers in manually scoring cells as a "CTC" or "no CTC". The results showed that, although all reviewers are trained according to the same guidelines, there is a major disagreement on a definition of a CTC. This motivates the use of automated methods. We compared the given scores to the answer of an expert panel and the answer of a Deep Learning network trained to separate CTCs from other cells. Remarkably, the Deep Learning network and the reviewers showed the highest agreement.

This motivated further investigations of Deep Learning networks for cell classification. We ultimately trained a convolutional neural network to classify cells into CTCs and four more cell classes that can be found in EpCAM enriched blood samples from cancer patients. We analyzed how network architectures based on autoencoders differ from standard architectures used for cell classification in terms of their encoding and interpretation of the image input data. The chosen network, in combination with advanced visualization techniques, allows to not only classify cells but to reveal new cell populations and subclasses of known populations. This paves the way to identify all cell populations in the fluorescent images and to investigate their implication for the outcome of patients.

The work presented in this thesis resulted in the development of an open-source software toolbox for CTC analysis called ACCEPT (**A**utomated **C**TC **C**lassification, **E**numeration and **P**heno**T**yping). The toolbox facilitates and automates the process of detecting and classifying CTCs and other cell populations in fluorescent images obtained from various microscopic systems and is now actively being used by several research groups in the field of CTC research.

Samenvatting

Kanker is wereldwijd een van de belangrijkste doodsoorzaken. Het begint met de ontwikkeling van een primaire tumor en kan zich uiteindelijk door het hele lichaam verspreiden, wat uiteindelijk de meeste kankergerelateerde sterfgevallen veroorzaakt. Tumorcellen die loskomen van de tumor en de bloedbaan binnendringen, worden circulerende tumorcellen (CTCs) genoemd. Het proces van CTCs die zich op een andere locatie in het lichaam nestelen en daar nieuwe tumoren vormen, wordt metastaseren genoemd. De hoeveelheid CTC in bloed is sterk gecorreleerd aan de overlevingsduur van kankerpatiënten en hun aantal kan worden gebruikt om therapie te volgen. Toch zijn CTCs zeer zeldzaam en het detecteren, karakteriseren en tellen in afbeeldingen van fluorescent gelabelde cellen is zeer uitdagend werk en wordt vaak handmatig gedaan. Dit proces is gevoelig voor fouten en gebruikersvoorkeuren. Dit proefschrift is daarom gericht op het ontwikkelen van modellen voor beeldanalyse en machinaal leren om CTCs automatisch en nauwkeurig te detecteren en classificeren en daarmee de toegevoegde waarde van geautomatiseerde beeldanalysemethoden aan te tonen.

CTCs en hun fluorescentiesignalen kunnen sterk variëren in grootte, vorm en signaalintensiteit. We hebben de aanwezigheid van verschillende afmetingen en intensiteiten in de fluorescentie-afbeeldingen geadresseerd door de ontwikkeling van een segmentatiemodel voor verschillende schalen. Het model is bijna parameteronafhankelijk en extraheerde de schaalinformatie op basis van een spectrale decompositie van het ingangssignaal met behulp van niet-lineaire diffusiemodellen. Dit leidde tot een nauwkeurige segmentatie van de cellen, gecombineerd met een clustering van de cellen op basis van hun grootte en intensiteit. We hebben het model verder aangepast tot een clustering van de cellen puur gebaseerd op grootte, die ook zeer zwakke fluorescentiesignalen uitstekend kan segmenteren.

Op basis van de segmentatie van alle cellen die in de afbeeldingen aanwezig zijn, kunnen we kwantificeerbare kenmerken per cel extraheren, zoals hun grootte of gemiddelde fluorescentie-intensiteit. Door deze geëxtraheerde kenmerken aan onderzoekers te geven die de fluorescente-expressie van CTCs als positief of negatief moeten beoordelen, kan de gebruikersbias sterk worden verminderd. Hierdoor kunnen de resultaten van meerdere onderzoekers, die voor de behandeling van de patiënt gebruikt worden, meer worden verenigd.

Verder hebben we de geëxtraheerde functies gebruikt om nieuwe celpopulaties te identificeren die ook in kankerpatiënten tot overexpressie worden gebracht, in vergelijking met gezonde proefpersonen. Het identificeren en begrijpen van deze celpopulaties heeft de potentie om de variatie tussen patiënten in ziekteprogressie en behandelingsrespons beter te begrijpen.

We analyseerden verder de consensus van meerdere reviewers in het handmatig scoren van cellen als een “CTC” of “geen CTC”. De resultaten lieten zien dat, hoewel alle reviewers volgens dezelfde richtlijnen zijn getraind, er grote onenigheid bestaat over een definitie van een CTC. Dit motiveert het gebruik van geautomatiseerde methoden. We hebben de scores van de reviewers vergeleken met het antwoord van een panel, bestaande uit experts in het gebied van CTCs, en met het antwoord van een Deep Learning netwerk, dat getraind is om CTCs te herkennen. Opmerkelijk genoeg toonden het Deep Learning netwerk en de reviewers de hoogste overeenkomst met elkaar.

Dit motiveerde het verdere onderzoek van Deep Learning netwerken voor classificatie van cellen. We hebben uiteindelijk een convolutioneel neuraal netwerk getraind om cellen in CTCs en nog vier andere celklassen te classificeren die gevonden kunnen worden in EpCAM-verrijkte bloedmonsters van kankerpatiënten. We hebben geanalyseerd hoe netwerkarchitecturen op basis van autoencoders verschillen van standaardarchitecturen die worden gebruikt voor classificatie van cellen, in termen van hun codering en interpretatie van de beeldinvoergegevens. Het gekozen netwerk, in combinatie met geavanceerde visualisatietechnieken, maakt het niet alleen mogelijk om cellen te classificeren, maar ook om nieuwe celpopulaties en subklassen van bekende populaties te identificeren. Dit maakt de weg vrij om alle celpopulaties in de afbeeldingen van fluorescente cellen te identificeren en hun implicatie voor de uitkomst van patiënten te onderzoeken.

Het werk gepresenteerd in dit proefschrift heeft geleid tot de ontwikkeling van een open-source software-toolbox voor CTC-analyse, genaamd ACCEPT (**A**utomated **C**TC **C**lassification, **E**numeration and **P**heno**T**yping). De toolbox vergemakkelijkt en automatiseert het proces van het detecteren en classificeren van CTCs en andere celpopulaties in fluorescentie-afbeeldingen verkregen van verschillende microscopische systemen en wordt nu actief gebruikt door verschillende onderzoeksgroepen op het gebied van CTC-onderzoek.

Publications

JOURNAL ARTICLES

1. *Multiscale segmentation via Bregman distances and nonlinear spectral analysis*
   Leonie L. Zeune, Guus van Dalum, Leon W.M.M. Terstappen, Stephan A. van Gils, Christoph Brune
   SIAM journal on imaging sciences 10.1 (2017): 111-146
2. *Combining Contrast Invariant L1 Data Fidelities with Nonlinear Spectral Image Decomposition*
   Leonie L. Zeune, Stephan A. van Gils, Leon W.M.M. Terstappen, Christoph Brune
   International Conference on Scale Space and Variational Methods in Computer Vision, Springer, Cham (2017): 80-93
3. *Quantifying HER-2 expression on Circulating Tumor Cells by ACCEPT*
   Leonie L. Zeune, Guus van Dalum, Charles Decraene, Charlotte Proudhon, Tanja Fehm, Hans Neubauer, Brigitte Rack, Marianna Alunni-Fabbroni, Leon W.M.M. Terstappen, Stephan A. van Gils, Christoph Brune
   PloS one 12.10 (2017): e0186562
4. *How to agree on a CTC - Evaluating the consensus in Circulating Tumor Cell scoring*
   Leonie L. Zeune*, Sanne de Wit*, A.M. Sofie Berghuis, Maarten J. IJzerman, Leon W.M.M. Terstappen, Christoph Brune
   Cytometry Part A 93.12 (2018): 1202-1206
   * Both authors contributed equally.
5. *Classification of Cells in CTC-Enriched Samples by Advanced Image Analysis*
   Sanne de Wit*, Leonie L. Zeune*, Christoph Brune, Guus van Dalum, T. Jeroen N. Hiltermann, Harry J.M. Groen, Leon W.M.M. Terstappen
   Cancers 10.10 (2018): 377
   * Both authors contributed equally.
6. *Deep Learning of Circulating Tumor Cells*
   Leonie L. Zeune, Yoeri Boink, Guus van Dalum, Afroditi Nanou, Sanne de Wit, Kiki C. Andree, Joost F. Swennenhuis, Christoph Brune, Leon W.M.M. Terstappen
   Submitted for publication (2018)

7. *Improving the CellSearch® system*
   Joost F. Swennenhuis, Guus van Dalum, Leonie L. Zeune, Leon W.M.M. Terstappen
   Expert review of molecular diagnostics 16.12 (2016): 1291-1305.
8. *Adaptive Classification of Arbitrary Activities Through Hidden Markov Modeling with Automated Optimal Initialization*
   Chris T. Baten, Thijs Tromper, Leonie L. Zeune
   Wearable Robotics: Challenges and Trends, Springer Int. Publishing (2017): 367-371
9. *Circulating tumor cells, tumor-derived extracellular vesicles and plasma cytokeratins in castration-resistant prostate cancer patients*
   Afroditi Nanou, Frank A.W. Coumans, Guus van Dalum, Leonie L. Zeune, David Dolling, Wendy Onstenk, Mateus Crespo, Mariane Sousa Fontes, Pasquale Rescigno, Gemma Fowler, Penny Flohr, Christoph Brune, Stefan Sleijfer, Johann S. de Bono, Leon W.M.M. Terstappen
   Oncotarget 9 (2018): 19283-19293
10. *Towards a real liquid biopsy in metastatic breast and prostate cancer: Diagnostic LeukApheresis increases CTC yields in a European prospective multi-center study (CTCTrap)*
    Kiki C. Andree, Anouk Mentink, Leonie L. Zeune, Leon W.M.M. Terstappen, Rui P. Neves, Nikolas H. Stoecklein, Christiane Driemel, Rita Lampignano, Liwen Yang, Hans Neubauer, Tanja Fehm, Johannes C. Fischer, Elisabetta Rossi, Mariangela Manicone, Umberto Basso, Piero Marson, Rita Zamarchi, Yohann Loriot, Valerie Lapierre, Vincent Faugeroux, Marianne Oulhen, Francoise Farace, Gemma Fowler, Mariane Sousa Fontes, Bernie Ebbs, Maryou Lambros, Mateus Crespo, Penny Flohr, Johann S. de Bono
    International Journal of Cancer 143.10 (2018): 2584-2591
11. *Single tube liquid biopsy for advanced non-small cell lung cancer*
    Sanne de Wit, Elisabetta Rossi, Sabrina Weber, Menno Tamminga, Mariangela Manicone, Joost F. Swennenhuis, Catharina G.M. Groothuis-Oudshoorn, Riccardo Vidotto, Antonella Facchinetti, Leonie L. Zeune, Ed Schuuring, Rita Zamarchi, T. Jeroen N. Hiltermann, Michael R. Speicher, Ellen Heitzer, Leon W.M.M. Terstappen, Harry J.M. Groen
    International Journal of Cancer (2019)
12. *A microwell array platform to print and measure biomolecules produced by single cells*
    Fikri Abali, Joska Broekmaat, Arjan G.J. Tibbe, Richard B.M. Schasfoort, Leonie L. Zeune, Leon W.M.M. Terstappen
    Submitted for publication (2018)

CONFERENCE CONTRIBUTIONS

Oral Presentations

1. *Mathematical Imaging and Inverse Problems for Analyzing Tumor Cells*
   Leonie L. Zeune, Guus van Dalum, Leon W.M.M. Terstappen, Stephan A. van Gils, Christoph Brune
   NDNS+ PhD days, April 23-24, 2015, Lunteren, The Netherlands
2. *ACCEPT - Automated CTC Classification, Enumeration and Phenotyping*
   Leonie L. Zeune, Christoph Brune, Guus van Dalum, Leon W.M.M. Terstappen
   CTCTrap Progress Meeting, September 8-9, 2015, Tallinn, Estonia
3. *Multiscale Cell Segmentation Using Spectral TV Analysis*
   Leonie L. Zeune, Christoph Brune, Guus van Dalum, Leon W.M.M. Terstappen
   1st Applied Mathematics Symposium, September 28-30, 2015, Muenster, Germany
4. *ACCEPT - Automated CTC Classification, Enumeration and PhenoTyping*
   Leonie L. Zeune, Guus van Dalum, Christoph Brune, Leon W.M.M. Terstappen
   2nd Cancer-ID General Assembly Meeting, March 21, 2016, Hamburg, Germany
5. *Multiscale Segmentation via Bregman Distances and Spectral TV Analysis*
   Leonie L. Zeune, Guus van Dalum, Leon W.M.M. Terstappen, Stephan A. van Gils, Christoph Brune
   SIAM Conference on Imaging Science, May 23-26, 2016, Albuquerque, USA
   SIAM Student Travel Award
6. *Multiscale Segmentation via Bregman Distances and Spectral TV Analysis*
   Leonie L. Zeune, Guus van Dalum, Leon W.M.M. Terstappen, Stephan A. van Gils, Christoph Brune
   NDNS+ Workshop, July 4-5, 2016, Enschede, The Netherlands
7. *ACCEPT - Automated CTC Classification, Enumeration and PhenoTyping*
   Leonie L. Zeune, Christoph Brune, Guus van Dalum, Leon W.M.M. Terstappen
   CTCTrap Progress Meeting, August 25-26, 2016, London, UK
8. *Contrast Invariant L1 Data Fidelities for Nonlinear Spectral Image Decomposition*
   Leonie L. Zeune, Guus van Dalum, Leon W.M.M. Terstappen, Stephan A. van Gils, Christoph Brune
   International Conference on Scale Space and Variational Methods in Computer Vision, June 4-8, 2017, Kolding, Denmark
9. *Standardization of CTC Classification using ACCEPT*
   Leonie L. Zeune, Guus van Dalum, Christoph Brune, Leon W.M.M. Terstappen
   Cancer-ID General Assembly Meeting, October 4-7, 2017, Rhodes, Greece
10. *Nonlinear Spectral Image Decomposition and its Application to Segmentation*
    Leonie L. Zeune, Guus van Dalum, Leon W.M.M. Terstappen, Stephan A. van Gils, Christoph Brune
    SIAM Conference on Imaging Science, June 5-8, 2018, Bologna, Italy

Poster Presentations & Abstracts

1. *Identifying Circulating Tumor Cells (CTCs) by Image Analysis*
   Leonie L. Zeune, Christoph Brune, Stephan A. van Gils, Guus van Dalum, Leon W.M.M. Terstappen
   MIRA Day, June 3, 2015, Enschede, The Netherlands
2. *Multiscale Segmentation of Tumor Cells Using Bregman Distances*
   Leonie L. Zeune, Christoph Brune, Stephan A. van Gils, Guus van Dalum, Leon W.M.M. Terstappen
   NDNS+ Workshop, June 22-23, 2015, Enschede, The Netherlands
3. *Identifying Circulating Tumor Cells (CTCs) by Image Analysis*
   Leonie L. Zeune, Christoph Brune, Stephan A. van Gils, Guus van Dalum, Leon W.M.M. Terstappen
   MIRA Day, November 3, 2016, Enschede, The Netherlands
   Best Poster Award
4. *Automated Identification of Circulating Tumor Cells by Image Analysis*
   Leonie L. Zeune, Guus van Dalum, Francois C. Bidard, Jean-Yves Pierga, Tanja Fehm, Hans Neubauer, Brigitte Rack, Marianna Alunni-Fabroni, Mateus Crespo, Johann S. de Bono, Leon W.M.M. Terstappen, Christoph Brune
   AACR conference, April 1-5, 2017, Washington D.C., USA
5. *Evaluating the Consensus in Circulating Tumor Cell Scoring*
   Leonie L. Zeune, Sanne de Wit, Guus van Dalum, Kiki C. Andree, Agustin Enciso Martinez, Joost F. Swennenhuis, Afroditi Nanou, Anouk Mentink-Leusink, Leon W.M.M. Terstappen, Leonie Majunke, Beate Zill, Marianna Alunni-Fabbroni, Rita Lampignano, Liwen Yang, Hans Neubauer, Tanja Fehm, Mariangela Manicone, Elisabetta Rossi, Rita Zamarchi, Francoise Farace, Johann S. De Bono, Christoph Brune
   ACTC Conference, October 4-7, 2017, Rhodes, Greece
6. *Evaluating the Consensus in Circulating Tumor Cell Scoring*
   Leonie L. Zeune, Sanne de Wit, A.M. Sofie Berghuis, Guus van Dalum, Maarten J. IJzerman, Christoph Brune, Leon W.M.M. Terstappen
   DCC-NET Retreat, November 3, 2017, Nettetal, Germany
7. *Deep Learning to identify Circulating Tumor Cells by ACCEPT*
   Leonie L. Zeune, Guus van Dalum, Afroditi Nanou, Sanne de Wit, Kiki C. Andree, Joost F. Swennenhuis, Leon W.M.M. Terstappen, Christoph Brune
   ISMRC Conference, May 3-5, 2018, Montpellier, France
   Best Poster Award
8. *Liquid biopsy in advanced NSCLC: EpCAM+ and EpCAM - circulating tumor cells, tumor derived extracellular vesicles and cell-free circulating tumor DNA*
   Sanne de Wit, Menno Tamminga, Ellen Heitzer, Michael R. Speicher, Ed Schuuring, T. Jeroen N. Hiltermann, Joost F. Swennenhuis, Leonie L. Zeune, Leon W.M.M. Terstappen, Harry J.M. Groen
   AACR conference, April 1-5, 2017, Washington D.C., USA

9. *EpCAM+ and EpCAM- circulating tumor cells in metastatic cancer patients: a multicenter study*
   Sanne de Wit, Mariangela Manicone, Elisabetta Rossi, Elisabeth K. Trapp, Rita Lampignano, Marianne Oulhen, Mateus Crespo, Leonie L. Zeune, Kiki C. Andree, Joost F. Swennenhuis, Riccardo Vidotto, Rita Zamarchi, Marianna Alunni-Fabbroni, Marie Tzschaschel, Brigitte Rack, Hans Neubauer, Tanja Fehm, Emeline Colomba, Francoise Farace, Penny Flohr, Johann S. de Bono, Leon W.M.M. Terstappen
   AACR conference, April 1-5, 2017, Washington D.C., USA
10. *Multicenter evaluation of technology platforms for the enumeration of circulating tumor cells*
    Sebastian Bender, Merlin V. Lütke-Eversloh, Rui P. Neves, Nikolas H. Stoecklein, Leon W.M.M. Terstappen, Barbara Baggiani, Martin H.D. Neumann, Thomas Krahn, Klaus Pantel, Thomas Schlange, Leonie L. Zeune
    AACR conference, April 1-5, 2017, Washington D.C., USA
11. *Circulating tumor cells, tumor derived extracellular vesicles and plasma cytokeratins in castration-resistant prostate cancer patients*
    Afroditi Nanou, Guus van Dalum, Leonie L. Zeune, Frank A.W. Coumans, Wendy Onstenk, Mateus Crespo, Stefan Sleijfer, Johann S. de Bono, Leon W.M.M. Terstappen
    ACTC Conference, October 4-7, 2017, Rhodes, Greece
12. *Single tube liquid biopsy for NSCLC*
    Sanne de Wit, Elisabetta Rossi, Menno Tamminga, Ellen Heitzer, Riccardo Vidotto, Mariangela Manicone, Leonie L. Zeune, Antonella Facchinetti, T. Jeroen N. Hiltermann, Ed Schuuring, Michael R. Speicher, Rita Zamarchi, Leon W.M.M. Terstappen, Harry J.M. Groen
    ACTC Conference, October 4-7, 2017, Rhodes, Greece
13. *Improving identification of cells enriched by CellSearch*
    Sanne de Wit, Leonie L. Zeune, Leon W.M.M. Terstappen
    ACTC Conference, October 4-7, 2017, Rhodes, Greece
14. *Classification of cell populations in CTC enriched samples by advanced image analysis*
    Sanne de Wit, Leonie L. Zeune, T. Jeroen N. Hiltermann, Harry J.M. Groen, Leon W.M.M. Terstappen
    AACR conference, April 14-18, 2018, Chicago, USA

Acknowledgments

Knowledge is in the end based on acknowledment.
Ludwig Wittgenstein

After 4.5 years of research, I finally reached the end of my thesis. Therefore, I want to take the opportunity to thank the people who contributed to this work and who supported and encouraged me during all the time.

A special thanks goes to my supervisors Leon Terstappen, Christoph Brune and Stephan van Gils for your support and guidance over the years. I am happy that Christoph motivated me to come to Twente and that Leon gave me the opportunity to work in this fascinating field of research. I am deeply grateful to both of you for all the valuable meetings and your input to my work and for encouraging me to pursue the research that I liked. I am grateful to Stephan for welcoming me in his group and for his kind support for me and my project.

I want to thank Guus van Dalum for mentoring me in the first years of my PhD and for his contribution in developing the ACCEPT toolbox. I am very grateful to Sanne de Wit for a lot of patience in testing the toolbox and finding almost every bug in the code. Thanks a lot for your enthusiasm and ideas throughout all of our shared projects.

For many uploaded datasets and feedback on the ACCEPT toolbox, I want to thank all collaborators from the CANCER-ID project. I want to express my gratitude to the members of my graduation committee – prof. Ozan Öktem, prof. Carola Schönlieb, prof. Klaus Pantel, prof. Nick Stoecklein, prof. Raymond Veldhuis and prof. Michel van Putten. Thanks a lot for agreeing to be part of my committee and reviewing my thesis.

I want to thank all people in the SACS group at EWI and in the MCBP group at TNW for providing a very open, friendly and inspiring atmosphere at work. I am grateful to Hil, Jurgen, Yoeri, Manu and Matthias from SACS for many nice discussions, group meetings and coffee breaks. A special thanks goes to my favorite roomie Afroditi, Sanne, Agustin, Kiki and Anouk for a great time at MCBP and some unforgettable conference trips! A word of thanks must also go to the secretaries Ingrid, Marielle and Linda for helping out with many organizational things and flight bookings.

For numerours help in organizing my defense party, I want to thank my paranymphs Sanne and Afroditi. I am very happy to have both of you on my side during that special day.

Finally, I want to thank my family and friends for supporting me, keeping my mind away from work and for making life so worth living. A very big hug goes to my family – my parents Arnhild and Rüdiger, Lisa, Pi, Stefan, Tomas, Emil, Henri and ongoing... Thanks for always supporting me, loving me and cheering me up when needed.

For all his love and support, I want to thank my husband Stefan. Thank you so much for all your patience, for always believing in me and encouraging me to achieve my goals, for being the best distraction from work I could imagine and for making me so incredibly happy every single day!

Appendix A

ACCEPT Manual

The **ACCEPT** Software, developed within the EU Cancer-ID project, is an open source image analysis package for the **A**utomated **C**TC **C**lassification, **E**numeration and **P**heno**T**yping. The current version contains

- the *CTC Marker Characterization* tool to quantify the expression of markers on pre-scored CTCs
- the *Full Detection* tool to detect and analyze all events in the images
- the *Candidate Selection* tool to semi-automatically classify events based on predefined gates.

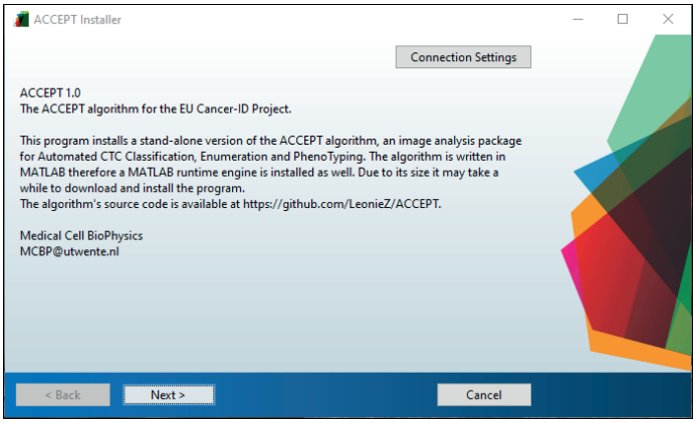
The different tools are called processors. A processor for fully-automated CTC classification is currently under development.

Installation of ACCEPT

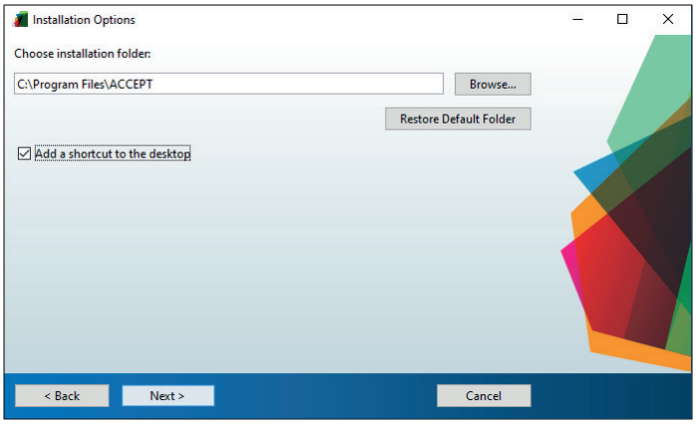
Installation of the stand-alone version To install the stand-alone version, download from *https://github.com/LeonieZ/ACCEPT/releases/download/1.2/ACCEPT_Windows.exe* the installer and double-click on it to start the installation. The start screen of the installation tool is shown in Figure A.1 (a). Click *Next* and on the following screen, shown in Figure A.1 (b), activate the tick-box to create a shortcut on the desktop (only for Windows users). Note that this is not the default option. If you use a Mac, the application will automatically be shown in your applications. In all following windows use the standard configuration and click *Next* or *Install* and wait until the installation is finished.

ACCEPT within MATLAB® If you have MATLAB® 2016b or a later version installed on your computer and do not want to use the ACCEPT stand-alone version, you can use ACCEPT directly within MATLAB®. From *https://github.com/LeonieZ/ACCEPT/archive/master.zip* down-

load the .zip folder and unpack it to your desired location. You start ACCEPT by running the ACCEPT.m script.



(a) Start Screen.

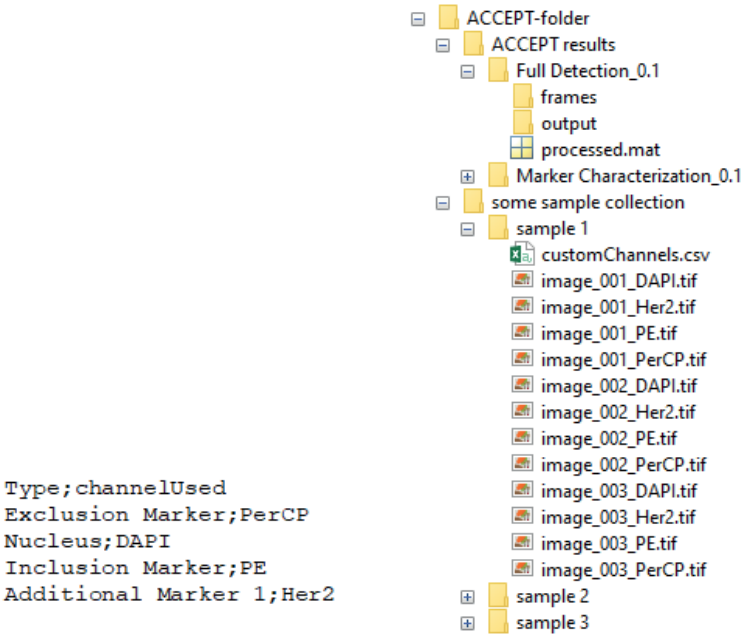


(b) Choose the installation folder.

Figure A.1: *Installation assistant of the stand-alone ACCEPT version.* Panel (a) shows the start screen of the installation assistant. In Panel (b) the user can choose the directory where the software will be installed and if a short-cut to the desktop should be generated.

Requirements for Image Data and Folder Structure

The images analyzed with ACCEPT are required to contain an *automated* scan of the full scanning area (no thumbnails of cells) and should *not* be stitched together. Moreover, they should be no compressed but raw .tiff files. In general, all scans that fulfill these conditions can be processed by ACCEPT. Yet, if your images are not archived from the CellSearch® system, a .csv file, specifying the names of the used channels, is required. The .csv file should be in the same folder as all your images and has to be named *customChannels.csv*. An ex-



(a) Content CSV file. (b) Required folder structure.

Figure A.2: *Required file and folder structure.* Panel (a) shows an example of a .csv file for a sample with four fluorescent marker where the specifier in the file names are *PerCP*, *DAPI*, *PE* and *Her2*. Panel (b) shows a sketch of the required folder structure.

ample with four different fluorescent channels is shown in Figure A.2 (a). For the execution of ACCEPT, two folders are required: one that contains the samples and a second one to save the outputs, the results produced by ACCEPT. Figure A.2 (b) shows a sketch of the folder structure. Each subfolder (*sample 1*, *sample 1*, ...) of the main input folder *some sample collection* is considered to contain an individual sample. The input folder, that should be specified, is the *some sample collection*-folder and *not* each sub-folder with an individual sample. The output folder cannot be placed inside the input folder but anywhere else on your computer. The shown folder structure will be automatically generated by ACCEPT. Once you organized your data in the described way, you can start using the software.

Main User Interface and Functions

The main user interface of the ACCEPT toolbox is visualized below. The main components are the chosen processors (Figure A.3 (1)), input and results folder, the sample list and the *Process* and *Visualize* buttons. More information about the different processors can be found in the next section. In the sample list, all samples found in the specified input folder are listed and are marked in red if they are not processed by ACCEPT before (with respect to the

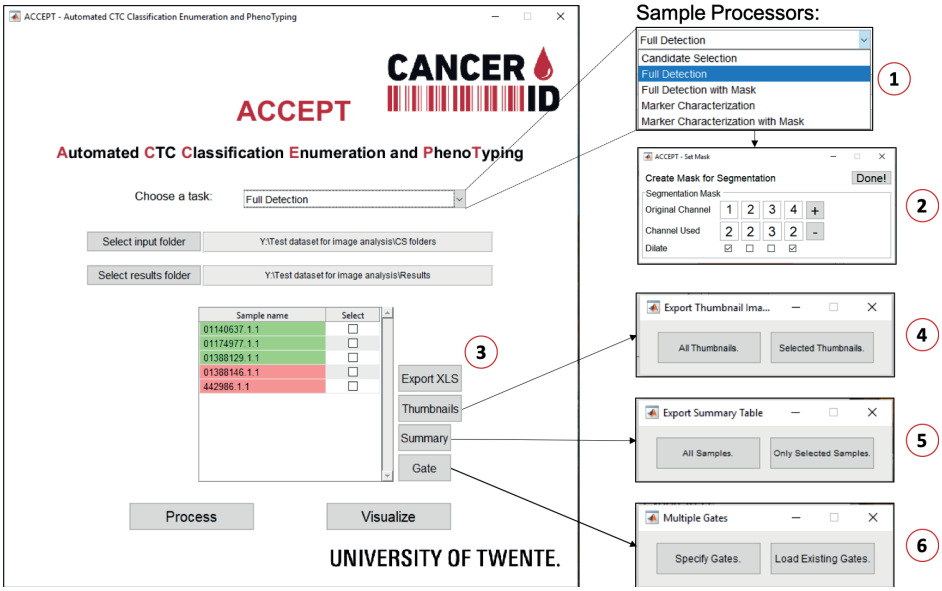


Figure A.3: Main User Interface of the ACCEPT toolbox and different functionalities of it.

specific processor and the output folder) and green if a result of that sample has already been saved in your results folder. Next to each sample name, there is a little tick-box for each sample that selects that sample for further processing and/or visualization. If you right-click in the column with the tick-boxes, all boxes will be ticked. With another right-click, all ticks will be removed again. If you want to visualize the results of a processed sample, you can select only one sample at a time. For all other tasks, you can select multiple samples at the same time. Apart from the two main buttons (*Process* and *Visualize*), there are four more buttons (3-6) as shown in Figure A.3:

- *Export XLS (3):* If you click on the *Export XLS* button, for every sample that is selected or, in case no sample is selected, for every sample that is processed, an excel file is stored in your results folder. The excel file contains the extracted measurements like mean intensity and size for every object found in the images of the associated sample.

- *Thumbnails (4):* If you click on the *Thumbnails* button, you have to specify if you want to export all thumbnails or only selected thumbnails that were classified before (for example the ones falling into a specific gate). Note that we restricted the number of saved thumbnails from a specified class to 200 images. If there are more cells that are classified in the class you selected before, 200 images are randomly picked. This should prevent that thousands of thumbnails are saved to your disk. The exported images are not optimized to be used in presentations or publications but to be processed by the thumbnail-based processors of the toolbox.

- *Summary (5):* If you click on the *Summary* button, you can first select, if you want a summary of all samples in the list or only the selected ones. You can then specify where you want to save the summary excel file. The created excel file will contain for every listed sample the number of events that fall into a certain gate if the gate has been applied to the specific sample before. If the gate has not been applied to some of the samples, the respective cell in the excel file will be empty. The table contains all gates found in at least one of the listed samples.
- *Gate (6):* If you click on the *Gate* button, you first have to select if you want to specify a new gate or load an existing gate (more details can be found in the *Design of Linear Gates* section). If you selected samples by ticking the checkbox, only the selected samples are gated; otherwise the gate is applied to all samples and samples that are not processed before will be processed with the selected processor.

Sample Processors

All available sample processors can be found under the *Choose a task* option in Figure A.3 and are described below.

Full Detection Processor The Full Detection processor is the processor to choose if you want to analyze *all objects, especially all fluorescently labeled cells* present in your images. If you are interested in other cell populations apart from CTCs that you scored before, this is the right sample processor. All objects in the images of the samples that are processed will be identified (using the algorithm described in Chapter 1). After detecting all objects, the following measurements are extracted for each object and each fluorescent channel:

1. size (in μm^2 (for CellSearch, otherwise in pixels))
2. eccentricity (circularity measurement between 0 (perfect circle) and 1 (line))
3. perimeter (in pixels)
4. mean intensity
5. max intensity
6. median intensity
7. standard deviation of the intensity
8. mass (sum of the intensity)
9. perimeter2area (P2A, circularity measurement, 1 for a perfect circle)
10. relative overlay with the signal in the nucleus channel

The processing takes roughly between 10–40 minutes per sample and depends on the number of cells found. To analyze the results and apply gates to specify cell classes, open the sample in the Sample Visualizer.

Marker Characterization Processor The aim of the Marker Characterization tool is to evaluate the expression of certain markers on *prescored CTCs*. That means, instead of loading all images, only small thumbnails of cells manually scored before as a CTC are loaded and evaluated. These cells can be either scored on the CellTracks Analyzer II® system or using a plugin that we developed for the ICY image analysis software (for more information see *https://www.utwente.nl/en/tnw/mcbp/protocolsandtools*). For each CTC that was loaded, the ten measurements per fluorescent channel that were mentioned before are extracted. Note that any object touching the border of the loaded thumbnails is excluded from further analysis and if multiple objects are found in one thumbnail, objects without any signal in the inclusion (third) fluorescent channel are removed from the analysis as well. Again, to analyze the results, open the sample visualizer by clicking on the *Visualize* button.

Full Detection with Mask Processor If you want to analyze all events in a sample but do not want to segment them in all fluorescent channels separately, you can use the Full Detection with Mask processor. In that case, the segmentation from one channel is copied to another channel. This can be very helpful if you have a fluorescent signal in one channel that is very difficult to analyze, for example if the signal is very dim or scattered. If you select this processor and click *Process*, a window will open where you have to specify for every fluorescent channel which channel should be used to copy the segmentation from (see Figure A.3 (2)). With the plus and minus signs, you can adjust the number of channels you used in your sample. The numbers in the upper row always remain as shown. In the second row (in the screenshot filled with 2 2 3 2) you can now specify the segmentation mask you want to use. If you fill in 1, 2, 3, 4 (so the first and the second row are the same) this processor does exactly the same as the Full Detection processor. In the example shown, the signal in channel 1 and 4 are not detected separately but the contour surrounding the signal in channel 2 is used in both channel 1 and channel 4. Note that for each channel, you have to select if you want to dilate the mask or not. If you select that tick-box, the mask from the *channel used* is copied but inflated by a few pixels. This is helpful, if you expect that the fluorescent signal in the *original channel* is smaller than the signal in the *channel used*. This is for example the case if you want to copy the contour surrounding the cell nucleus (channel 2) to a channel that should stain the whole cell body. If you fill in a 0 in the second row, the corresponding channel will be excluded from further analysis.

Marker Characterization with Mask Processor Similarly to the relation between the Full Detection and Full Detection with Mask processor, this processor does the same analysis as the Marker Characterization processor but allows you to copy the segmentation from one channel to another.

Candidate Selection Processor The Candidate Selection processor is based on the Full Detection processor, but is followed by a semi-automated classification of the found objects based on manually defined gates. Once you click on the *Process* button, a window to specify your gate appears. More information on using the gates can be found in the *Design of Linear*

Gates section (Figure A.5). Once you specified or loaded your gate, click *Done!* and the processing starts. The results are visualized in the Sample Visualizer, the result of the gating is loaded using the *Load Selection* button (Figure A.4 (5)).

Sample Visualizer

After processing, the results can be visualized in the *Sample Visualizer* shown in Figure A.4. The visualization tool has the following components and functions:

Sample Information: In the Sample Information box (top left) all sample information retrieved from the data files accompanying the images of a sample is contained. Which information can be read and displayed here, depends on the type of sample.

Overview Image: Next to the extracted information, a stitched overview image of all the images in the sample is shown. Below the display is a box indicating the channel that is illustrated. By clicking on the box, you can alter the fluorescent channel. The overview image can be used to assess the number of cells and the quality of the sample. If you see a systematic background error in all images (half of the image is yellow, half blue) or if you see that the sample is very full, the segmentation quality might suffer and you have to be careful when analyzing the results.

Thumbnail Gallery: In the Thumbnail *Cell Gallery* all objects that were evaluated are shown and the images are linked to the scatter plots on the right. The left thumbnail is an overlay of the first three fluorescent channels. Red is used for the first fluorescent channel (in this case CD45), blue for the second channel (DNA) and green for the third channel (CK). The white scale bar in the overlay represents 10 pixels. In this example the pixel size is 0.64 μm (retrieved from the sample information). Therefore, the bar has a size of 6.4 μm. The red lines in the thumbnails show the contour of the object as identified by the image analysis algorithm. You can look at all objects by using the scroll function at the right of the thumbnails or use the spacebar on your keyboard to make jumps of five. To change the overlay from a three channel overlay to a two channel overlay, deactivate the *3 Channel Overlay* box (see (4) in Figure A.4). After deactivation, the overlay image shows only the overlay of the second and third channel in the colors known from the CellSearch® system. Note that the overlay image is scaled from the minimum to the maximum value within the evaluated signal (indicated by red contour) while the other thumbnails that display only one fluorescent channel show the full intensity range and therefore often appear dimmer. If you make a right-click on the images showing a single channel, an enlarged version of the image pops up and here the signal is always scaled. Moreover, the measurements extracted in this fluorescent channel are listed. If you want to add a thumbnail to your selection, left-click on the overlay image and a blue frame appears and the corresponding point in the scatter plot is highlighted as well.

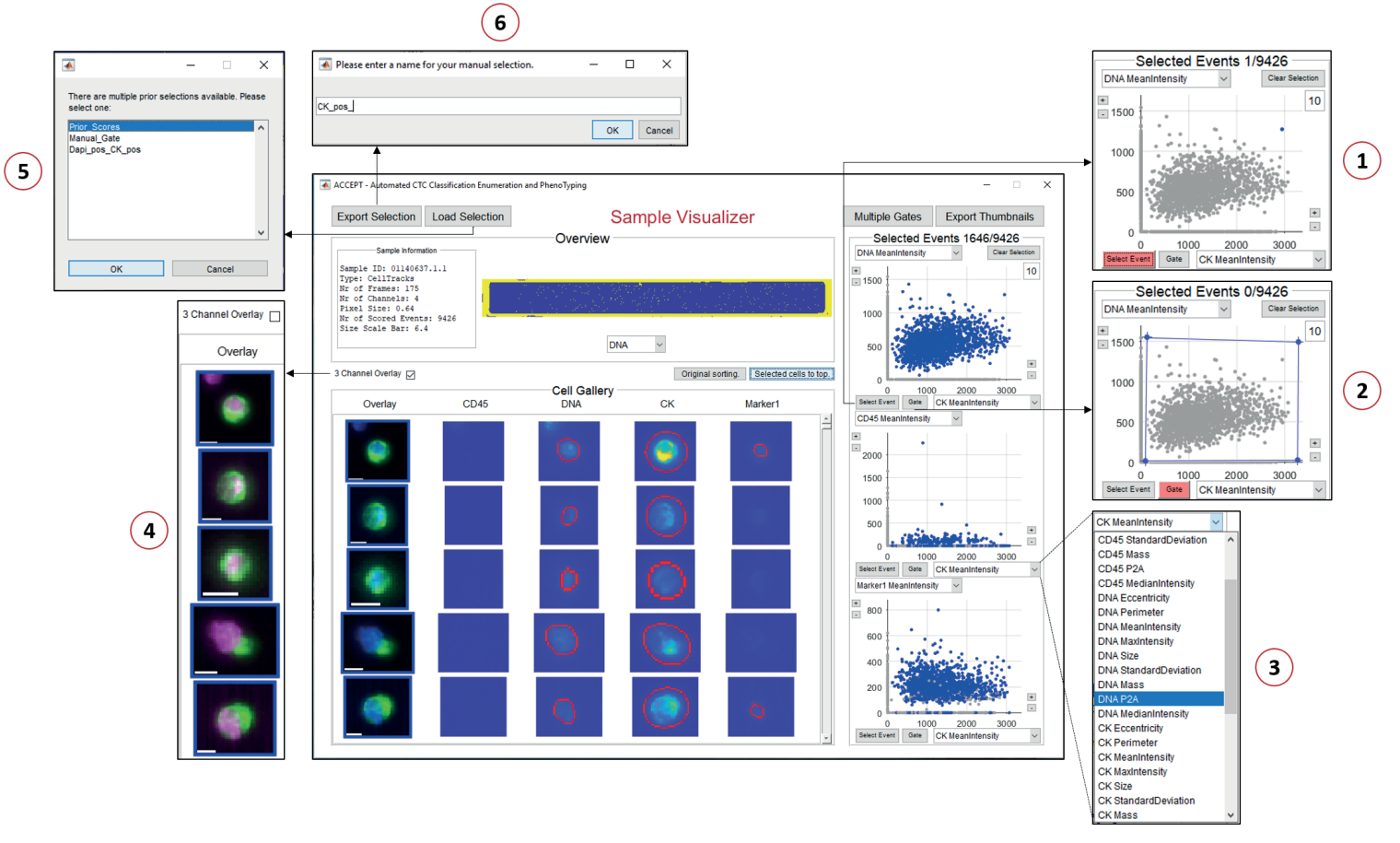


Figure A.4: *ACCEPT interface to visualize the results after processing.* The main window is depicted in the middle and surrounded by different functionalities of the toolbox labeled from (1) to (6). The different functionalities are described in the text.

Scatter Plots: On the right are three scatter plots. You can change the displayed features of all parameters by clicking on the box next to the x or y-axis (see (3) in Figure A.4). For each channel you can select the Area, Eccentricity, Perimeter, Mean Intensity, Median Intensity, Max Intensity, Standard Deviation of the Intensity, Mass (Sum of the Intensity), Perimeter2Area (a circularity measurement) and the relative overlay with the second fluorescent channel (nucleus). The dots in the scatter plot represent all objects identified in the sample. By clicking on the + and – buttons next to the x or y-axis you can increase or decrease the scale at which the objects are displayed. By changing the number 10 in the top right corner of the top scatter plot, you can alter the size of the dots. Above the three scatter plots, you see the total number of events found (in this example 9426 objects) and how many of them are currently selected.

Selection of an Event: By pushing the *Select Event* button under each scatter plot ((1) in Figure A.4) you can click on a dot in the scatter plot which turns blue and the thumbnail associated with this object appears at the third position in the thumbnail gallery at the left and is highlighted by a blue contour drawn around the overlay thumbnail. The object also appears blue in the two other scatter plots.

Gating in the Scatterplots: Next to the *Select Event* button is the *Gate* button ((2) in Figure A.4). Pushing this button allows you to select all points falling in a region of the scatter plot that you can manually define. You start to label the region by clicking in the scatter plot and mark every corner of your region by moving the mouse to that location and click again. Close the region by returning to the starting point (a blue open circle appears when reaching the starting point). Once a gate is specified, all objects that fall into the specified gate are depicted in blue. All thumbnails associated to a point in the selected region are highlighted by a blue contour around the overlay thumbnail.

Gating with a Gate Set: For the simultaneous use of more features from the measured parameters, you can use the *Multiple Gates* feature shown in Figure A.5. After a click on the button in the top right corner, the display shown in Figure A.3 (6) pops up. You can either select to create a new gate (*Specify Gates*) or use an already existing gate (*Load Existing Gates*). In case you choose to specify a gate, the display shown in Figure A.5 pops up. For more details see the section *Design of Linear Gates*.

Resorting of the Gallery: By clicking on the *Selected cells to top* button, the thumbnail gallery is resorted so that all selected thumbnails are shown at the top of the thumbnail gallery followed by all other thumbnails. To return to the original sorting click the *Original sorting* button. If you update your selection (by deselecting or adding events), update the thumbnail gallery by clicking the *Selected cells to top* button again.

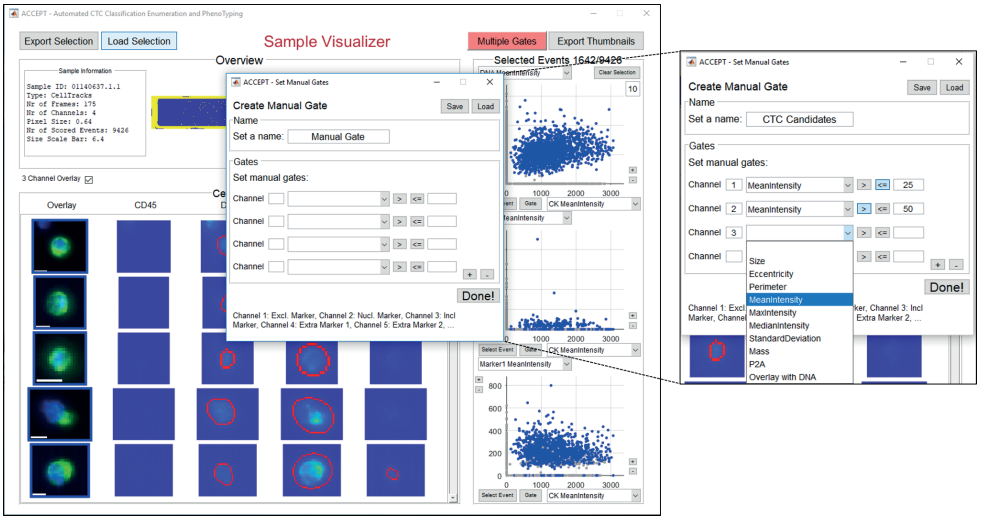


Figure A.5: *ACCEPT interface for the creation of linear gates.* On the left, the starting window of the interface is shown when called via the Sample Visualizer. An example on how to create a gate set is shown in the right.

Exportation and Loading of a Selection: If you want to save a cell selection, you can click on the *Export Selection button* and specify a name for your selection (see (6) in Figure A.4). Note that the name is not allowed to contain any gaps or special characters. Previously stored selections or gates can be loaded by the *Load Selection* button. If multiple selections are stored for the sample that is shown, you can select which one you want to visualize. Panel (5) in Figure A.4 shows an example; in this case three selections (named *Prior Scores*, *Manual Gates* and *Dapi_pos_CK_pos*) were stored.

Exportation of Thumbnails: In case you want to store all or only the selected thumbnails as .tiff files on your hard drive, click *Export Thumbnails*. The saved thumbnails can be found in the results folder specified in the main user interface. Note that the thumbnails are optimized for further processing with the toolbox, not for presentations or publications.

Design of Linear Gates

In order to search for certain types of objects (like CTCs, Leukocytes, ...) you can design sets of linear gates that you can apply to the extracted measurements and visualize the results afterwards. You can call the window to design and apply the gates at multiple locations in the toolbox (Main User Interface, Sample Visualizer and Candidate Selection Processor). The window that will open is shown in Figure A.5. On the top, you can set the name of the gate, again with no gaps and no special characters. With the plus and minus signs at the bottom, you can adjust the number of parameters that you want to use to restrict the region where

the events fall in. For every lower and upper bound, you have to specify in which fluorescent channel you want to apply the gate (1 for CD45, 2 for DAPI, 3 for CK, ...) and choose in the drop-down list which parameter you want to gate. Afterwards, specify if it is a lower bound (click "$>$") or a higher bound ("$\leq$") and then enter the threshold value. Once you designed your gate, you can either directly apply it by clicking *Done!* or save your gate first.

Save a Gate: If you plan to use a gate more often, it is most convenient to save it and reload it once you want to apply the gate to a new set of samples. To do so, click *Save*. There is a default place to save the gates (the toolbox will always start looking for gates at this place) but you can also browse to wherever you want to save the gate and select *Save*.

Load a Gate: If you want to load a gate that you already saved (for example to make minor adjustments before applying it or to apply it as it is), click the *Load* button and browse to your saved gates and select one. It is then visualized in the window for designing the gates.

Download and Further Information

ACCEPT can be downloaded from *https://github.com/LeonieZ/ACCEPT*. A description of specific use-cases can be found under *https://github.com/LeonieZ/ACCEPT/wiki*. For further information about the software contact accept-mcbp-tnw@utwente.nl.