

Habilitation thesis reviewer's report

Masaryk University

Faculty Faculty of Informatics, MU

Habilitation field Informatics

Applicant RNDr. David Svoboda, PhD

Unit Faculty of Informatics

Habilitation thesis Virtual Cell Imaging (methods and techniques)

Reviewer Prof. Carolina Wahlby

Unit Centre for Image Analysis, Dept. of Information Technology,
Uppsala University, Uppsala, Sweden

Reviewer's report

The topic of the habilitation thesis of Dr. Svoboda is concerned with simulation of image data mimicking real microscopy images. Simulation is an important component of any technology development, but perhaps even more so in fields where creation of 'ground truth' or known answers, is extremely time-consuming, prone to bias, and variable among visual annotators. Microscopy in life science is such a field. Microscopy has been one of the most important techniques for understanding biology and medicine ever since the first microscopes were invented in the 17th century. Today, automated microscopes produce enormous amount of image data with multiple spatial, spectral- and time-resolution. Digital image processing techniques that can find objects of interest and quantify phenotypic changes in the objects observed are thus urgently needed. Many techniques exist, and many new approaches are continuously published. This is where the work of Dr. Svoboda becomes so important; without data connected to 'ground truth' it is very difficult to compare, evaluate and validate new emerging techniques.

Dr. Svoboda's work on simulation of image data stems from his own development of image analysis techniques, and his own need to validate the performance of his algorithms. One of his first contributions in image analysis includes improvements to the use of deformable models for tissue reconstruction using dual simplex meshes. Furthermore, he has introduced k-NN based classification and feature extraction algorithms customized for cell analysis, and carefully analysed per-feature discrimination power and cell-level classification accuracy. He has also made significant contributions to texture analysis for describing fluorescence signal distributions in microscopy data.

In his paper in CVIU from 2014 Dr. Svoboda explores a number of edge detection strategies from a statistical point of view, and shows his deep knowledge in careful algorithm evaluation in relation to scale and noise complexity. This work has also been continued in parallel to the work on synthetic images, with focus on real-world image analysis problems within the field of microscopy related to important biomedical questions. One example is his work on histone methylation published in Physiological Research in 2007; it demonstrates Dr. Svoboda's ability to collaborate across disciplines and create algorithms that are applicable.

Biomedical image data is often degraded by the point spread function of the imaging system, and deconvolution algorithms are often needed to correct for such degradation effects.

However, traditional deconvolution algorithms are slow, especially for large 3D image sets. This is another field where Dr. Svoboda has made significant contributions: He has developed computational strategies for speeding up the algorithms by parallel computations on multiple GPUs. Furthermore, he has developed efficient computational methods for more broadly needed convolutional filters, decomposing the convolutional step to reduce computational time as well as memory consumption.

Dr. Svoboda has made several solid contributions to the area image simulation research. Among them is a careful analysis of the distribution and organization of cells in 3D tissue samples, and the simulation of tissue. Another is the investigation and simulation of the distribution of chromatin inside cell nuclei, where Dr Svoboda proposed and analysed the effect of many of the different phenomena (biological/chemical as well as physical) affecting the quality of microscopy image data. This research not only leads to better data simulation, but also to a better understanding of the degradation phenomena as such. Another important contribution is Dr. Svoboda's algorithms for generation of 3D cell nuclei, published in Cytometry in 2009. It has been cited in more than 20 papers and used to benchmark a number of novel cell segmentation algorithms. His more recent (2016) simulations on vascular networks hold great potential for contributing to the understanding of chemical signalling and formation of vascular networks, e.g. in the process of tumour formation.

One of the most important contributions when it comes to simulation of time-lapse data is Dr. Svoboda's benchmark dataset for comparison of cell tracking algorithms, published in Bioinformatics in 2014, together with a number of leading scientists in the field of bio-image informatics. This paper has already led to more than 40 citations, and represents a thorough comparison of the state-of-the-art of the field. Image analysis challenges are growing in popularity, benchmarking algorithms against one another. They have made it possible for algorithm developers from all over the world to compete and compare their approaches on equal terms, without the bias of lab reputation or contacts.

Dr. Svoboda has consistently provided to the bio-image informatics community by providing his tools (simulation tools as well as algorithms) free and open source via the web, making them available for researchers and students world-wide, and strongly contributing to the field and the important aspects of reproducible science.

For the future, it would have been interesting to extend this line of research with a comparison of the presented simulation techniques to the new emerging field of adversarial nets for generating synthetic images by simultaneously training two models: one generative model that captures the data distribution, and a discriminative model that estimates the probability that a sample came from the real training data rather than the synthetically generated image, see Goodfellow et al, NIPS 2014

<https://papers.nips.cc/paper/5423-generative-adversarial-nets>. Such data generation is currently limited as model parameters are often difficult to extract, but future combined approaches may help us better model the complex networks and processes of biology.

Summarising, Dr. Svoboda has made a large number of relevant contributions in the area of synthetic image generation and cell image analysis. He has clearly demonstrated his expertise in this field and shown his ability to lead a well-motivated and productive research program. Given these achievements, I am happy to recommend acceptance of his habilitation thesis.

Reviewer's questions for the habilitation thesis defence

1. One of the fastest growing trends in digital image processing and analysis is machine learning based approaches, such as deep convolutional neural networks, for image segmentation and classification. Such methods require large amounts of annotated image data for training. How do you think simulated data, such as the data produced by your methods, can impact (or is already having an impact) on the research field?
2. When simulating data, visual similarity and computational similarity to real microscopy data may differ. How can you make design choices that evaluate these aspects in a fair way, or should the two aspects of simulation be regarded as two different design goals?
3. What are the most important next developments needed in the field of simulation of biomedical image data, and why?

Conclusion

The habilitation thesis submitted by David Svoboda entitled "*Virtual Cell Imaging (methods and techniques)*" **meets** the requirements applicable to habilitation theses in the field of Informatics.

Uppsala, May 22, 2017