Generating labels to solve inverse problems in computational biophotonics

Anant S. Vemuri[a], Sebastian J. Wirkert[a], Hannes Kenngott[b], Sara Moccia[c], Michael Götz[c], Benjamin Mayer[b], Patrick Mietkowski[b], Martin Wagner[b], Peter Sauer[e], Neil T. Clancy[f,g], Daniel S. Elson[g], and Lena Maier-Hein[a]

[a]Div. of Computer Assisted Medical Interventions, DKFZ, Heidelberg
[b]Dept. for General, Visceral and Transplantation Surgery, Heidelberg University Hospital, Germany
[c]Dept. of Advanced Robotics, Istituto Italiano di Tecnologia, Genoa, Italy
[d]Div. of Medical Image Computing, DKFZ, Heidelberg
[e]Department of Gastroenterology, Toxicology and Infectious Diseases, University Hospital Heidelberg, Heidelberg, Germany
[f]Univ. College London, UK
[g]Hamlyn Centre for Robotic Surgery, Imperial College London, UK

Abstract

Machine learning has achieved breakthrough successes in various problem domains in recent years. However, in certain fields, lack of labeled data can be regarded a major bottleneck. An example is the field of multispectral imaging (MSI), a sub-field of biophotonics that finds widespread application in disease diagnosis, tissue assessment and image-guided surgery [1]. It is well-known that tissue reflectance captured by a MSI camera encode physiological properties including oxygenation and perfusion, among other properties which are not visible to the naked eye. Yet, decoding this information is an inverse problem, which has not yet been adequately solved with model-based methods, which suffer from too simplistic base assumptions and long run times.

One key aim of the European Research Council (ERC) project COMBIOSCOPY is to leverage the potential of machine learning techniques in the context of biophotonics in general and MSI in particular.

As acquisition of reference measurements for training algorithms is practically not feasible (due the lack of a gold standard technique), our concept relies heavily on simulated data generated with Monte Carlo methods [2]. Yet, instead of using tissue-specific optical parameters (which are hard to obtain), we propose using a generalized parameter set that encompasses most of the tissue types in the body, to build large amounts of simulated labeled data. We then efficiently combine the simulated labeled data with few unlabeled "target specific" *in vivo* samples using the concept of transfer learning [3].

According to porcine trials as well as first in human studies, our methodology allows us to recover important optical parameters, such as oxygenation, with high accuracy and in real-time under different clinical conditions (including different anatomy) without having explicit *in vivo* reference labels in the training data. Our general approach to solving inverse problems with machine learning could thus become a valuable tool in the field of medical image computing and computer assisted interventions.

References

[1] Baowei Fei Guolan Lu. Medical hyperspectral imaging: a review. *Journal of Biomedical Optics*, 19:19 – 19 – 24, 2014.

[2] Sebastian J. Wirkert, Hannes Kenngott, Benjamin Mayer, Patrick Mietkowski, Martin Wagner, Peter Sauer, Neil T. Clancy, Daniel S. Elson, and Lena Maier-Hein. Robust near real-time estimation of

physiological parameters from megapixel multispectral images with inverse monte carlo and random forest regression. *International Journal of Computer Assisted Radiology and Surgery*, 11(6):909–917, Jun 2016.

[3] Sebastian J. Wirkert, Anant S. Vemuri, Hannes G. Kenngott, Sara Moccia, Michael Götz, Benjamin F. B. Mayer, Klaus H. Maier-Hein, Daniel S. Elson, and Lena Maier-Hein. Physiological parameter estimation from multispectral images unleashed. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2017*, pages 134–141. Springer International Publishing, 2017.