PhD project proposal – CIM

Development of Multimodal AI-supported Image Data Analysis Methods for Improved Cancer Diagnostics

Purpose and aims

Deep learning and Deep Convolutional Neural Networks have revolutionized the field of Computer Vision, with applications ranging from face recognition to self driving cars. This revolution is now propagating into the fields of medical and biomedical imaging, mobilizing the power of deep learning for radically improved health care.

Oral cavity cancer (OCC) and oropharyngeal cancers (OPC) are among the most common malignancies, leading to more than 220 000 premature deaths each year [1]. Incidence in Sweden is around 1200 cases with around 350 deaths every year. Early diagnosis is of highest importance for treatment and survival, however the symptoms of OCC and OPC are diffuse and often mild, leading to delayed diagnosis and significantly worse prognosis.

In an international multi-center effort, we are, for a number of years, working towards AI-supported screening for OCC and OPC based on brush samples. The project connects top researchers and end-users in Sweden – Uppsala University (UU), Karolinska University Hospital (KS), Falun Hospital, Public Dental Health Services (FTV) in Stockholm, Blekinge and Dalarna – and several cancer centres in India. Towards our goal, we have developed a deep learning based diagnostic support system [2] which demonstrates feasibility of large scale usage. We now focus our effort on increasing the specificity of the system, towards highly accurate predictions of cancer type and patient prognosis, which would greatly benefit clinical treatment of cancer patients.

Cancer is a complex disease; its different causes and types have a strong impact on patient treatment and prognosis. This has motivated an intensified search for factors with prognostic relevance. To improve understanding of the disease and its progression, as well as enabling reliable early detection, **this PhD project** will explore and develop techniques for multimodal information fusion, utilizing combinations of several imaging modalities to maximize information gain. The project combines the power of modern deep learning techniques [3] with novel distance measures between images [4], to create **new methods for efficient fusion of multimodal image data** – with the primary aim to improve cancer diagnostics, while contributing to, and building on, general theoretical development of the field.

This is an interdisciplinary project which connects theoretical method development, based on advancements in mathematics, information theory, and computer science, with practical requirements of medical applications.

The power of multimodal imaging

Modern imaging techniques reveal a variety of properties of a specimen – morphology, chemistry, dynamics, function – however, most often only one such property at a time. For full understanding, different techniques have to be combined. We will harvest from the availability of microscopy techniques which provide complementary information about a specimen and combine several modalities towards improved oral cancer diagnostics. Complementing our existing brightfield (BF) data, Second Harmonic Generation (SHG) imaging has demonstrated relevance in tissue-based cancer detection [5] while Fluorescence microscopy of PAP stained samples has shown usefulness for studying the oral microbiome [6], strongly correlating with oral cancer. A very promising result [7] reveals spatial differentiation of the fluorescence distributions of the EA-50 PAP-stain between tumor cells and normal cells. We have initiated acquisition of images by brightfield, fluorescence, SHG, and electron microscopy (EM), to maximally provide complementary information about the specimen. Fusing the heterogeneous information about the imaged specimen, we aim to increase specificity, improve prognosis prediction, and facilitate improved understanding of the mechanisms driving the cancer progression.

Project plan

Success in utilization of multimodal information heavily relies on the quality of **image registration**, as well as approaches for **deep information fusion** – both will be developed and evaluated in this project. We will focus on few-shot, self-supervised, and unsupervised learning based approaches to multimodal image data analysis, to avoid the need for a large number of manually aligned image pairs, as well as to overcome the challenge of severe modality imbalance (the amount of bright field images will highly exceed the alternative modalities, the latter ones focusing on the positive detections).

Perfect alignment (registration) of the highly heterogeneous views is essential for information fusion. This requires efficient methods for multimodal image registration, in general an ill-posed and very difficult task. Existing image registration methods are restrictive w.r.t. the size of disalignment, and are often prohibitively slow in modern imaging scenarios. Their generality and applicability to new modalities are very limited.

Mutual information (MI) has long been the first choice of a similarity measure for multimodal image registration. It is also used in recent approaches for representation learning [8]. MI measures the statistical dependence between the intensities of the two signals, however without including contextual information. Being a global measure, its local estimation as well as optimization, are challenging tasks. Recent efforts are put in its analysis and estimation in the context of deep metric learning [9]; we will contribute to this effort, in addition to exploring ways to extend the context-awareness of this measure. We will build on theoretical properties of MI, integrating them with the excellent complementary properties of our proposed α-SMD distance measure [4], towards creating improved and computationally efficient similarity measures for large multimodal images.

We aim to further develop our recent approach to Cross-modality representation learning [3], enabling fast alignment of heterogeneous data. The approach involves learning joint cross-modal representations of the images and then optimizing (mono-modal) similarity between these representations, thereby transforming the multi-modal registration problem into a (simpler) mono-modal one, see Figure 1.

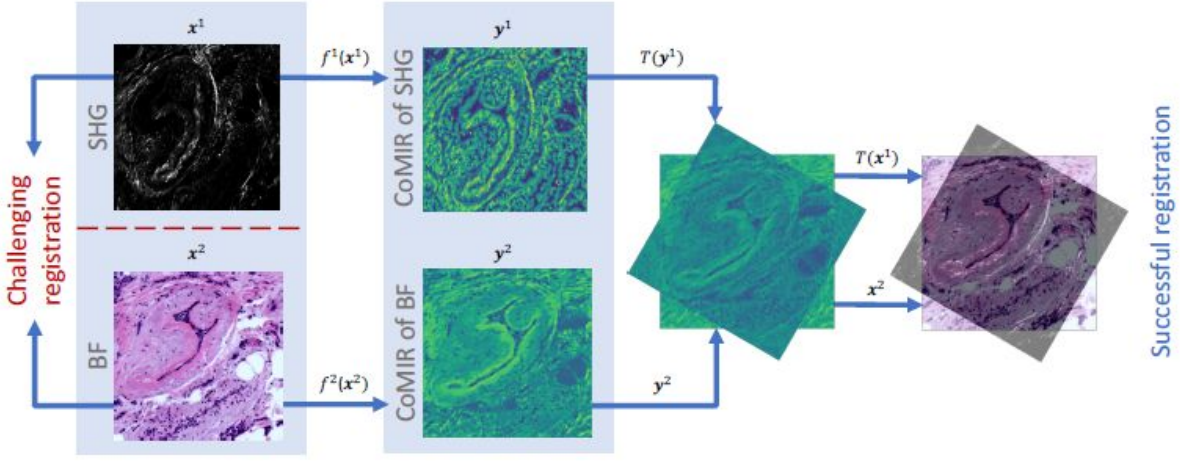

Figure 1: Using CoMIRs, different modalities such as BF and SHG can be registered with monomodal approaches based on alpha-AMD or SIFT.

The second project goal, efficient and general deep learning-based multimodal data fusion, is an open problem and a subject of active research [10,11]. Despite being a fairly common use case, the combination of few-shot learning and multimodal fusion is still in its infancy. We plan to develop few-shot and deep image translation approaches for efficient information fusion also when not all modalities are available – EM and SHG are expensive techniques which will not be available for large scale screening. We will explore joint representation learning approaches [3,11,12] and adversarial cross-modal data generation [13]. One approach, inspired by [14], to achieve this is to develop strategies to assist the monomodal DL system by attention transfer learning.

Host institution

The PhD candidate will be part of the MIDA group at the Division of Visual Information and Interaction, Department of Information Technology, Uppsala University.

The ideal candidate has a MSc in Computer Science, Applied Mathematics, Machine Learning, or related field with a broad mathematical knowledge as well as good programming skills. The components to be studied build on a number of mathematical techniques and will require good command of the related areas; central are information theory, mathematical optimization and probability theory, while scientific computing and differential geometry have prominent roles as well. Experience and/or interest in the medical sciences is beneficial.

The PhD position is jointly funded by the Centre of Interdisciplinary Mathematics (50%) and the Division of Visual Information and Interaction (50%), Dept. of Information Technology, Uppsala University.

Main advisor

Dr. Joakim Lindblad (PI, theoretical development, deep learning for image processing and analysis)
Division of Visual Information and Interaction, Dept of Information Technology, Uppsala University
e-mail: joakim.lindblad@it.uu.se Webpage: http://www.cb.uu.se/~joakim/

Co-advisor

MD, Dr. Kristina Edman (research strategist, sample acquisition, domain expertise in oral cancer)
Centrum för klinisk forskning Dalarna, Region Dalarna and Dept of Surgical Sciences, Uppsala University
e-mail: kristina.edman@regiondalarna.se

The team combines unique expertise for the task, where previous and ongoing work on AI-supported oral cancer diagnostics demonstrates well functional collaboration. Our active involvement in the H2020 COST Action COMULIS on multimodal imaging and analysis in life-sciences provides connection to the latest developments. Clinical expertise and involvement in day-to-day practice within the team ensures usefulness of the end results. Existing data lead to zero startup time. We are collaborating with the BioVis facility at UU for the imaging.

References

[1] Bray, F., et al. "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries." CA: a cancer journal for clinicians 68.6: 394-424, 2018.

[2] Lu, J., N. Sladoje, C. Runow Stark, E. Darai Ramqvist, J-M. Hirsch, J. Lindblad. A Deep Learning based Pipeline for Efficient Oral Cancer Screening on Whole Slide Images. LNCS 12132, pp 249-261, 2020.

[3] Pielawski, N., E. Wetzer, J. Öfverstedt, J. Lu, C. Wählby, J. Lindblad, N. Sladoje. CoMIR: Contrastive Multimodal Image Representation for Registration. NeurIPS, 2020.

[4] Öfverstedt, J., Lindblad, J., Sladoje, N. Fast and Robust Symmetric Image Registration Based on Distances Combining Intensity and Spatial Information. IEEE Trans. on Image Processing, 27(7) pp. 3584-3597, 2019.

[5] Perry, S.W., Burke, R.M. & Brown, E.B. Two-Photon and Second Harmonic Microscopy in Clinical and Translational Cancer Research. Ann Biomed Eng 40, 277–291, 2012.

[6] Lunawat, P.P. et al. Detection of acid fast bacilli in saliva using papanicolaou stain induced fluorescence method versus fluorochrome staining: An evaluative study. Journal of international oral health: JIOH, 7(7), p.115, 2015.

[7] Atyaoui, M., Dimassi, W., Tounsi, N., Jaidan, N.E. and Ezzaouia, H. Fluorescence spectroscopy and imaging to improve diagnosis of normal and tumoral cytological pancreatic cells. Pathology-Research and Practice, 209(1), pp.1-5, 2013.

[8] Bachman et al. Learning Representations by Maximizing Mutual Information Across Views. NeurIPS 2019.

[9] Tschannen et al. On Mutual Information Maximization for Representation Learning. ICLR 2019.

[10] Ramachandram, D. and Taylor, G.W. Deep multimodal learning: A survey on recent advances and trends. IEEE Signal Processing Magazine, 34(6), pp.96-108, 2017.

[11] Guo, W., Wang, J. and Wang, S. Deep multimodal representation learning: A survey. IEEE Access, 7, pp.63373-63394, 2019.

[12] Wetzer, E., N. Pielawski, J. Öfverstedt, J. Lu, J. Lindblad, ..., N. Sladoje. Cross-modal Representation Learning for Efficient Registration of Multiphoton and Brightfield Microscopy Images of Skin Tissue. 4th NEUBIAS Conference, France, 2020.

[13] Pahde, F., O. Ostapenko, P. J. Hnichen, T. Klein and M. Nabi, "Self-Paced Adversarial Training for Multimodal Few-Shot Learning," 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 218-226, 2019.

[14] Zhang, S. et al. MSAFusionNet: Multiple Subspace Attention Based Deep Multi-modal Fusion Network. In International Workshop on Machine Learning in Medical Imaging, pp.54-62, 2019.