Multimodal Contrastive Representation Learning - 3D CoMIR

Project proposal for the course 'Project in Computational Science'/Scientific Computing', Fall 2021, in Image Analysis and Machine Learning at the Centre for Image Analysis (CBA) at the Division for Visual Information and Interaction, Vi2.

Background

Multimodal images refer to images produced by multiple imaging techniques, such as different sensors. Combining the information of different imaging modalities offers complimentary information about the properties of the imaged specimen. Often these modalities need to be captured by different machines, which requires that the resulting images need to be registered in order to map the corresponding signals to each other. This can be a very challenging task due to the varying appearance of the specimen in different sensors.
We have recently developed a method which uses contrastive learning to find representations of both modalities, such that the images of different modalities are mapped into the same representational space. The learnt representations are abstract and very similar with respect to a selected similarity measure. These representations are referred to as CoMIRs [1]. There are requirements which these representations need to fulfil for the downstream task of registration, such as rotational equivariance [2] or intensity similarity; these requirements can be enforced through modifications of the contrastive loss, which is based on InfoNCE [3,4]. Finally, in this abstract space, common methods for monomodal registration (intensity-based as well as feature-based) can be used to align the corresponding images. The transformation aligning CoMIRs found this way can be applied to the original images, which overcomes the problem of multimodal registration, see Fig. 1.

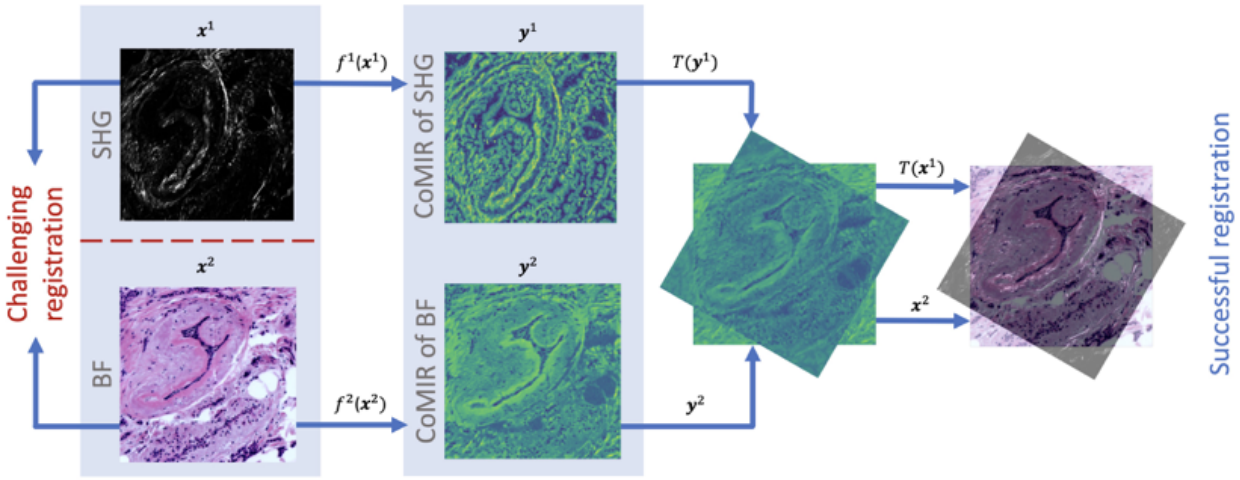

Fig.1: The transformation found for the CoMIRs is applied to the original modalities to achieve multimodal registration.

Aim

As the use of CoMIRs has shown great potential in registration applications in histology (e.g. registration of brightfield microscopy images and second harmonic generation images), the aim of this project is to extend the current representational learning and registration framework from 2D to 3D. Currently, in the 2D case, two U-Nets, one per modality, are trained without sharing weights, to find representations for the respective input, solely by being connected through the contrastive loss function. This setup has to be adjusted to process 3D image data, which includes the adaptation of the network architecture, and likely the way the data is processed, to meet the memory constraints imposed by the available (and targeted) hardware. Furthermore, the theoretical foundation to enforce equivariant feature maps needs to be extended to 3D in formulation and implementation.

The resulting method shall be implemented, tuned, and evaluated on a multimodal, biomedical dataset of scientific interest, and compared to at least one other suitable state of the art multimodal registration method.

Most of the available code (for two modalities of 2D images) is implemented in python, but evaluation methods for registration may rely on other languages or software packages.

Prerequisites

Well versed in advanced/scientific python programming; basic understanding of the underlying theory (similarity measures, equivariance/invariance); and experience with Deep Learning implementations (e.g., taken course “Deep Learning for Image Analysis).

References

[1] **Pielawski, Wetzer, et al: CoMIR: Contrastive Multimodal Image Representation for Registration.** In Proc. of NeurIPS 2020,
Paper online at:
https://proceedings.neurips.cc/paper/2020/hash/d6428eecbe0f7dff83fc607c5044b2b9-Abstract.html
Video presentation at:
https://slideslive.com/38937317/comir-contrastive-multimodal-image-representation-for-registration?ref=speaker-44922-latest
Poster at:
https://www.it.uu.se/research/visual_information_and_interaction/research/mida/NeurIPS2020.pdf
and
https://www.it.uu.se/research/visual_information_and_interaction/research/mida/ComulisElisabeth.pdf
Code available at: https://github.com/MIDA-group/CoMIR
[2] For an overview of CNN related methods which achieve equivariance, see **Lafarge et al.: Roto-translation equivariant convolutional networks: Application to histopathology image analysis**, Medical Image Analysis Feb. 2021
Paper online at: https://www.sciencedirect.com/science/article/pii/S1361841520302139#fig0003
[3]] **Hjelm et al.: Learning deep representations by mutual information estimation and maximization.** ICLR 2019,
Paper online at: https://openreview.net/forum?id=Bklr3j0cKX
[4] **Tschannen et al.: On Mutual Information Maximization for Representation Learning,** ICLR 2020,
Paper and Video online at: https://iclr.cc/virtual_2020/poster_rkxoh24FPH.html

Supervisors & Contact:

Nataša Sladoje, Professor: natasa.sladoje@it.uu.se
Elisabeth Wetzer, PhD student: elisabeth.wetzer@it.uu.se