Contrastive Learning for Equivariant Multimodal Image Representations

Elisabeth Wetzer[1], Nicolas Pielawski[1], Eva Breznik[1], Johan Öfverstedt[1], Jiahao Lu[2], Carolina Wählby[1], Joakim Lindblad[1], Natasa Sladoje[1]
[1]Uppsala University, Sweden; [2]University of Copenhagen, Denmark

Introduction

- Combining information of multiple modalities for one specimen can shed light on properties not detectable by only one modality as they can provide complementary details.
- Multimodal Registration can be extremely challenging if the appearance or signal expression density differs greatly between the modalities, as is the case for brightfield (BF) microscopy and second harmonic generation (SHG).
- We have developed a contrastive learning method based on InfoNCE [2] to learn representations from different modalities, called CoMIRs [1], which are visually similar.
- These image-like, dense representations can be succesfully registered by monomodal rigid registration methods, e.g. α-AMD (intensity-based, [3]) or using SIFT (feature-based, [4]).
- Top-10 cross-modal image retrieval success using CoMIRs for reverse image search is 65%, doubling the performance of direct cross-modal retrieval of the multimodal images.
- No data-specific information is incorporated in the learning, i.e. the method is modality independent and can be applied to other imaging modalities than BF and SHG.

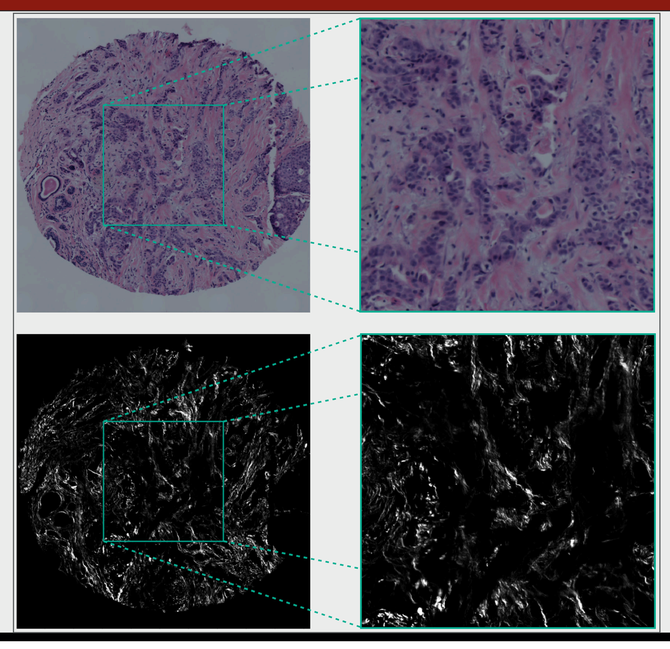


Fig. 1: 834x834px Patches to be registered and retrieved (available at [5]), cut from TMA cores captured by BF and SHG. In our performance evaluation random rotations by ± 30° and random translations by ±100px were applied to the patches.

Contrastive Learning

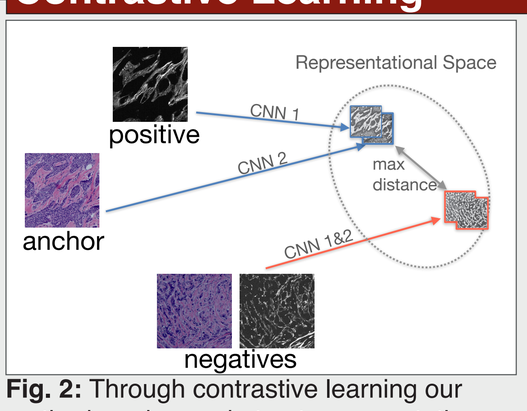


Fig. 2: Through contrastive learning our method produces abstract representations which are very similar for all input modalities.

- A randomly cropped patch in one modality serves as an anchor. Its corresponding patch in the other modality acts as a positive. Any other patch of any modality serves as a negative.
- Two CNNs, sharing no weights, only connected by the loss function, learn dense representations by maximizing the distance between the anchor and the negatives, as well as minimizing the distance between the anchor and the positive.

CoMIRs

- We require certain properties of the representations, such as rotational equivariance and similar intensities, which can be realized through the loss function without any additional hyperparameters.
- The appearance of CoMIRs depends on the choice of similarity function (critic); MSE yields the best results.
- The number of channels for the CoMIRs can be chosen; single channel CoMIRs expedite registration.

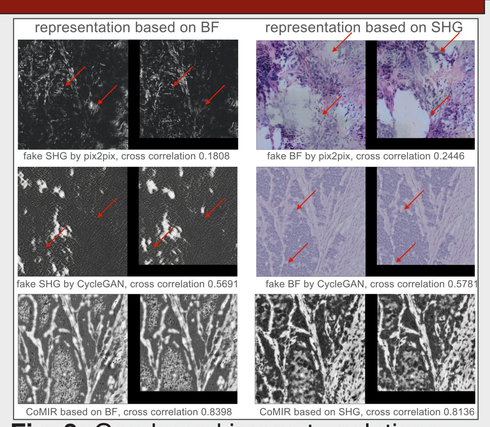


Fig. 3: Gan-based image translations, unlike CoMIRs, lack rotational equivariant properties.

Loss Modification for Rotation Equivariance

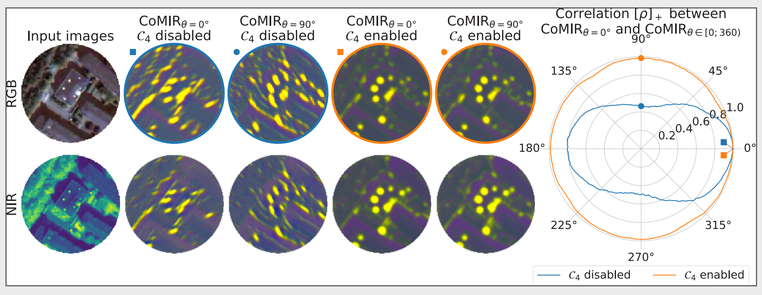


Fig. 4: Rotation Equivariance is achieved for any angle by commuting the CNN training with rotations of the C4 group, here shown on an arial image dataset.

- For an image pair (x^1,x^2), its CoMIRs (y^1,y^2), and a critic h, the loss based on InfoNCE [2] is given by

$$\mathcal{L}_{\boldsymbol{\theta}}(\mathcal{D}) = -\frac{1}{n}\sum_{i=1}^{n}\left(\log\frac{e^{h(\boldsymbol{y}_i^1,\boldsymbol{y}_i^2)/\tau}}{e^{h(\boldsymbol{y}_i^1,\boldsymbol{y}_i^2)/\tau}+\sum_{\boldsymbol{y}_j^1,\boldsymbol{y}_j^2\in\mathcal{D}_{neg}} e^{h(\boldsymbol{y}_j^1,\boldsymbol{y}_j^2)/\tau}}\right)$$

- Modification of the critic which allows for rotational equivariance commutes the CNN f_θ with actions T_i, T_i' of the C4 group (rotations by multiples of 90 degree).

$$h(\boldsymbol{y}_j^1,\boldsymbol{y}_j^2) \to h\left(T_1'\left(f_{\boldsymbol{\theta}_1}\left(T_1(\boldsymbol{x}_i^1)\right)\right), T_2'\left(f_{\boldsymbol{\theta}_2}\left(T_2(\boldsymbol{x}_i^2)\right)\right)\right)$$

Downstream Tasks

- **Registration:** CoMIRs can be registered by common, monomodal methods based on their intensities (e.g. α-AMD [3]) or by feature based methods (e.g. SIFT [4]).
- **Retrieval:** A Bag-of-Words [BoW,6] based on invariant feature extractors (SIFT/SURF [7]) and cosine similarity for matching can be used.

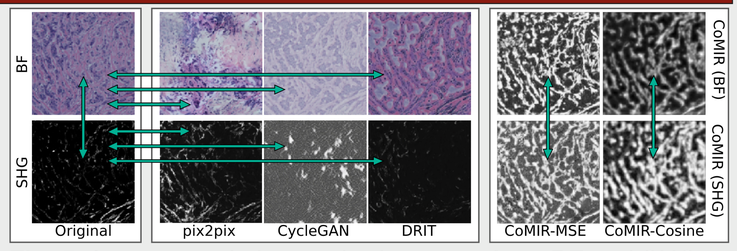


Fig. 5: Image translation methods to transform BF and SHG into one common modality. Arrows indicate resulting pairs for registration.

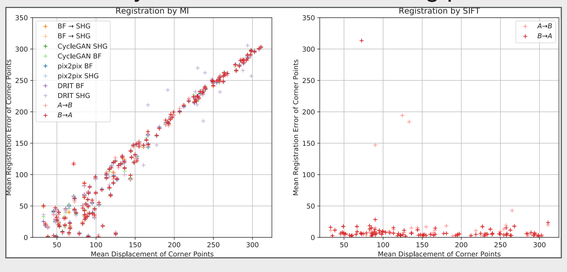
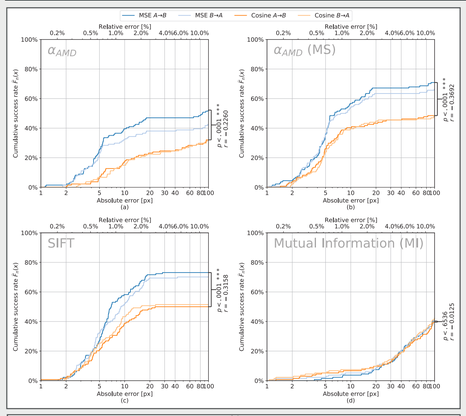


Fig. 6: MI fails to detect global extrema if the initial displacement is too large.

Results and Conclusions



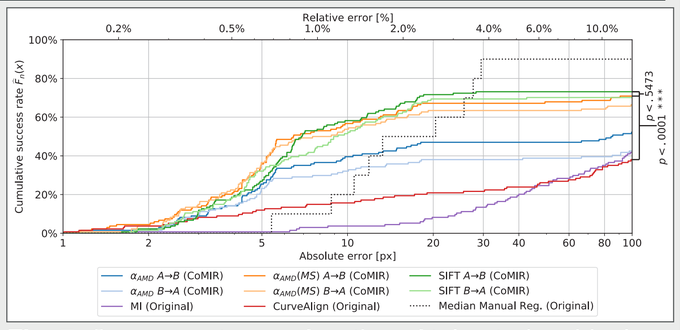


Fig. 7: Success rate as a function of tolerated residual error of registration of the CoMIRs based on MSE.

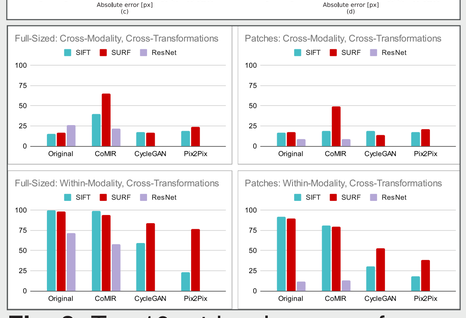


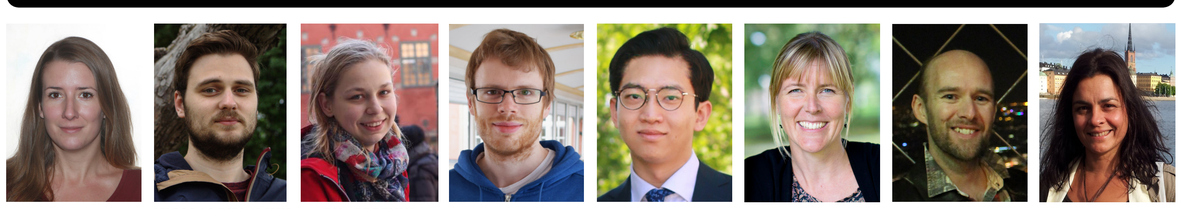
Fig. 8: Top-10 retrieval success for reverse image search shows the importance of equivariant representations and invariant feature extractors to handle transformations and multimodality.

- CoMIRs extract shared content in mutlimodal images and enable multimodal registration and retrieval by reducing the problem to a monomodal one.
- CoMIRs combined with monomodal intensity- and feature-based registration methods significantly outperform registration by mutual information and data-specific SotA.
- CoMIRs are suitable for cross-modal reverse image search when combined with invariant fearture extractors in a BoW.

References and Code

Code available at https://github.com/MIDA-group/CoMIR

[1] Pielawski, Wetzer et al.: CoMIR: Contrastive Multimodal Image Representation for Registration. NeurIPS, 2020
[2] Hjelm et al.: Learning deep representations by mutual information estimation and maximization. ICLR 2019
[3] Öfverstedt et al.: Fast and robust symmetric image registration based on distances combining intensity and spatial information. Trans. on Img Proc, 2019
[4] Lowe et al.: Object recognition from local scale-invariant features. ICCV 1999
[5] Eliceiri et al.: Multimodal Biomedical Dataset for Evaluating Registration Methods (patches from TMA Cores). 10.5281/zenodo.3874362, 2020
[6] Philbin et al.: Object retrieval with large vocabularies and fast spatial matching. CVPR 2007
[7] Bay et al.: SURF: Speeded Up Robust Features. CVIU, 2008



ELISABETH.WETZER@IT.UU.SE
NICOLAS.PIELAWSKI@IT.UU.SE
EVA.BREZNIK@IT.UU.SE
JOHAN.OFVERSTEDT@IT.UU.SE
JIAHAO.LU@DI.KU.DK
CAROLINA.WAHLBY@IT.UU.SE
JOAKIM.LINDBLAD@IT.UU.SE
NATASA.SLADOJE@IT.UU.SE

erc
Centre for Image Analysis Uppsala University
VINNOVA Sweden's Innovation Agency
WASP WALLENBERG AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM