Registration of Multimodal Microscopy Images using CoMIRs
Learned Structural Image Representations

Elisabeth Wetzer¹, Nicolas Pielawski¹, Johan Öfverstedt¹, Jiahao Lu², Carolina Wählby¹, Joakim Lindblad¹, Natasa Sladoje²
¹Uppsala University, Sweden; ²University 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 signals.
- Multimodal Registration can be extremely challenging if the appearance or signal expression density differs greatly between the modalities, as is the case for brightfield microscopy (BF) and second harmonic generation (SHG).
- We have developed contrastive learning based on InfoNCE [2] to learn representations from different modalities, called CoMIRs [1], which are visually similar.
- These image-like, dense representations can be successfully registered by monomodal rigid registration methods, e.g. a-AMD (intensity-based, [3]) or using SIFT (feature-based, [4]).
- 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.
- Very little aligned training data is required, for modalities which share sufficient structural similarities, the required aligned training data can be as little as one image pair.

Contrastive Learning
- 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 can be registered by common monomodal methods based on their intensities (e.g. a-AMD [3]) or by feature-based methods (e.g. using SIFT [4]).
- A transformation found for CoMIRs can be applied to original modalities and solve the multimodal registration.

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; MSE yielded the best results.
- The number of channels for the CoMIRs can be chosen; single channel CoMIRs expedite registration.

Competing Methods
- CurveAlign: registers BF and SHG, using modality specific information and mutual information (MI) [6].
- GAN-based image Translation methods: pix2pix, CycleGAN, DRIT. The resulting “fake” modalities do not qualify for intensity- or feature based monomodal registration methods.

Results and Conclusions
- CoMIRs extract shared content in multimodal images and enable multimodal registration by reducing the problem to a monomodal one.
- CoMIRs combined with monomodal intensity- and feature-based registration methods significantly outperform multimodal registration by MI as well as a state-of-the-art data-specific approach. Best results were obtained using SIFT.
- CoMIRs contain more valuable information than GAN generated images obtained by image-to-image translation from one modality to the other.
- Feature-based registration does not depend on the initial displacement of the images as is the case for MI-based approaches.

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