CoMIR: Contrastive Multimodal Image Representation for Registration

Nicolas Pielawski, **Elisabeth Wetzer**, Johan Öfverstedt, Jiahao Lu,
Carolina Wählby, Joakim Lindblad, Nataša Sladoje

NEURAL INFORMATION PROCESSING SYSTEMS

Method

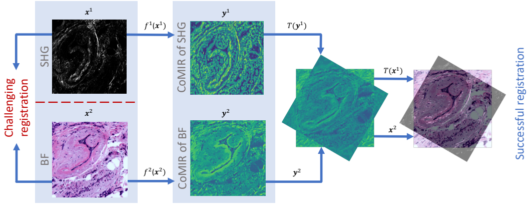

Figure 1: Using CoMIRs, different modalities such as BF and SHG can be registered with monomodal approaches based on alpha-AMD or SIFT.

We show that (1) contrastive learning of aligned pairs of images can produce representations, referred to as **CoMIRs**, that reduce complex (or even non-feasible), rigid multimodal registration to much simpler mono-modal registration; (2) the proposed **CoMIRs** are rotation equivariant due to a modification of a commonly used contrastive loss based on InfoNCE. Performance of the learned representations is evaluated on a very challenging task of multimodal registration of **bright-field (BF)** and **second-harmonic generation imaging (SHG)** by enabling monomodal registration methods on the CoMIRs.

- The proposed method trains a set of neural network on aligned image pairs of different modalities
- Requires **very little training data** (for some datasets as little as one image is enough)
- Learns image-like, dense representations named **CoMIRs**
- Representations have equivariant properties without additional hyperparameters
- **Enforces C4 equivariance** during training to achieve equivariance beyond multiples of 90°
- Evaluated on two datasets: **an aerial dataset** of Zurich (RGB images and near-infrared) and **a biomedical dataset** (BF and SHG microscopy images)

The modification of the critic in the InfoNCE loss (see equation 1) ensures equivariant properties of the representations. It is model-independent and does not require any architecture modifications, nor any additional hyperparameter tuning.

Rotation Equivariance

iN5GlPWFZ_Q

Watch on YouTube

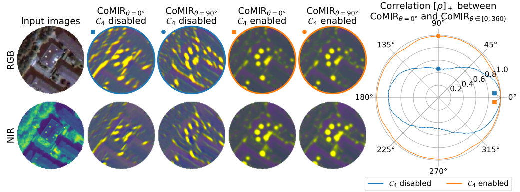

Figure 2: The rotation equivariance demonstrated on the aerial dataset.

To ensure rotation equivariance of the CoMIRs, we optimize

$$\underset{\theta_1,\theta_2}{\operatorname{argmax}}\ \mathbf{sim}\Big(T_1'\big(f_{\theta_1}(T_1(x^1))\big), T_2'\big(f_{\theta_2}(T_2(x^2))\big)\Big) \tag{1}$$

where $T_i, T_i' \in \mathcal{T} = \mathcal{C}_4$ the finite, cyclic, symmetry group of multiples of 90° rotations, x^i the input image in modality i and f_{θ_i} the network processing the respective modality.

Stability across training

The CoMIRs generated by our method can vary drastically across runs. We show that the embeddings can be reproduced when the neural network is **initialized with a fixed seed** and the same training image, even if the training consists of randomly chosen batches of images. Figure 3 shows that the CoMIRs of the same image produced by 50 neural networks are consistent over training.

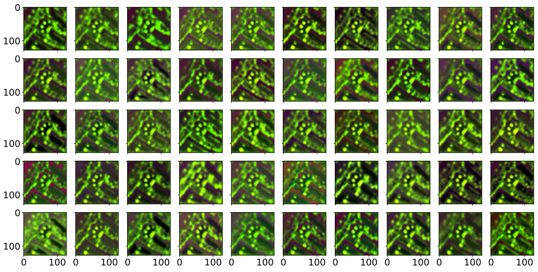

Figure 3: CoMIRs of 50 neural networks on the same input image (same weight init.).

Results

zpcgnqcQgqM

Watch on YouTube

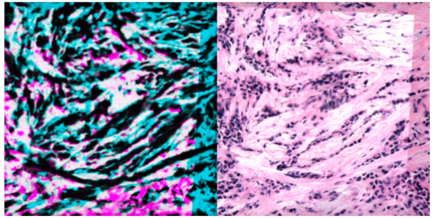

Figure 4: Example of registration (left: differences, right: real images overlapped)

We achieved an accuracy of 80% (successfully registered images with an error of less than 100px), which doubled the previous SOTA success rate.

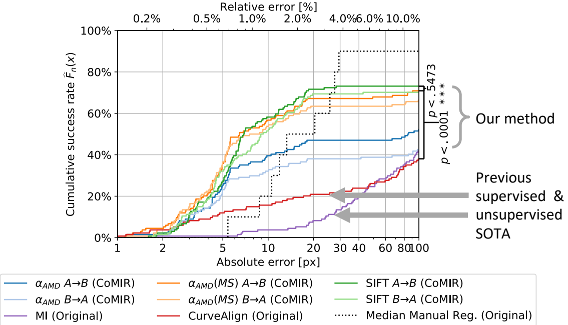

Figure 5: Registration accuracy on the biomedical dataset.

nicolas.pielawski@it.uu.se

elisabeth.wetzer@it.uu.se

CoMIR on Github
github.com/MIDA-group/CoMIR