UNIVERSITY OF COPENHAGEN

# Deep Feature based Image Registration

**Steffen Czolbe, Oswin Krause, Aasa Feragen**
per.sc@di.ku.dk, oswin.krause@di.ku.dk, afhar@dtu.dk

## Summary


- Image registration algorithms minimize a similarity-based loss function
- Existing pixel-based metrics like mean squared error ($\mathrm{MSE}$) and normalized cross-correlation ($\mathrm{NCC}$) have trouble dealing with noise, intensity differences, and mismatches in image anatomy.
- Our Method $\mathrm{DeepSim}$ measures the alignment of semantic feature embeddings. Registration models trained with our metric converge faster and achieve a higher registration accuracy compared to $\mathrm{NCC}$ on noisy datasets.


## Method

Given a set of feature-extracting functions $F^l : \mathbb{R}^{\Omega \times C} \to \mathbb{R}^{\Omega_l \times C_l}$ for $L$ layers of abstraction, we define a cosine-similarity based metric

$$\mathrm{DeepSim}(\mathbf{I}, \mathbf{J}) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{|\Omega_l|} \sum_{\mathbf{p} \in \Omega_l} \frac{\left\langle F^l_{\mathbf{p}}(\mathbf{I}), F^l_{\mathbf{p}}(\mathbf{J}) \right\rangle}{\|F^l_{\mathbf{p}}(\mathbf{I})\| \|F^l_{\mathbf{p}}(\mathbf{J})\|} \quad . \qquad (1)$$

$F^l_{\mathbf{p}}(\cdot)$ denotes a vector over $C_l$ output channels of the feature-extraction function at pixel $\mathbf{p}$. The neighborhood of the pixel is considered in the metric, as we compose $F^l$ of convolutional filters with increasing receptive area of the composition.

## Experiments

The feature-extractors $F^l(\cdot)$ need to extract semantic features relevant to the images. First, we use a supplementary segmentation task to train a model with the desired properties of the feature extractor. In a second step, we use the convolutional filters of the segmentation model's encoder to extract descriptive features for the similarity metric during registration. For comparison, we train registration networks with multiple baseline functions.

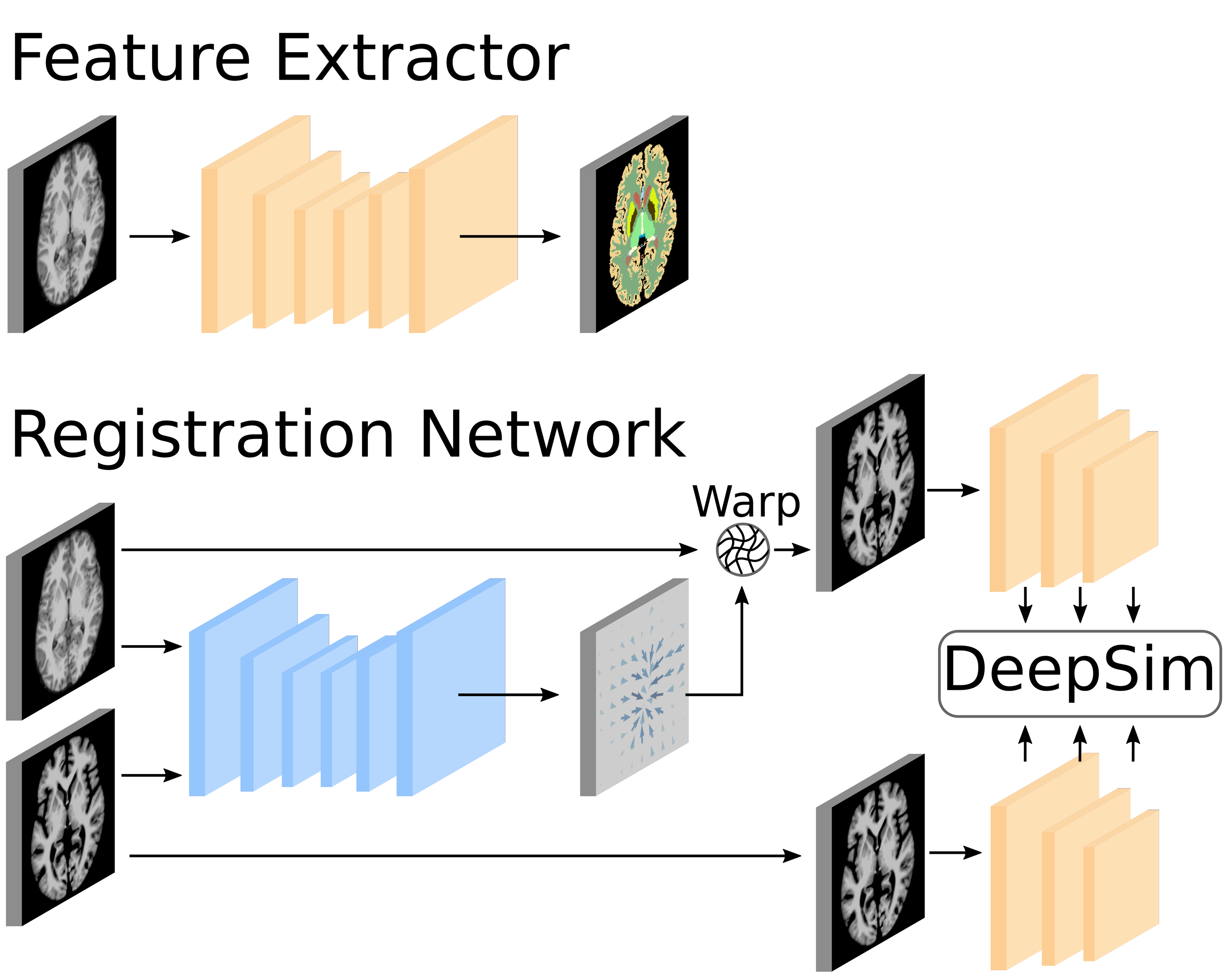


**Figure 1:** We use identical U-Net architectures for segmentation and registration. The fixed encoder of the segmentation model is used as the feature extractor for training the registration model.

## Data

We evaluate our method on three publicly available datasets of different modalities and dimensionality:

- Platelet-EM, a noisy, multi-class 3D serial block-face scanning electron microscopy image of human blood cells. We register neighboring 2D slices.
- PhC-U373, a high contrast 2D+time dataset of Glioblastoma-astrocytoma cell movements from the ISBR cell tracing challenge. We register along the time dimension.
- T1-weighted Brain-MRI scans, neatly preprocessed using Freesurfer. We combine 4400 3D scans from the ABIDE and OASIS3 studies and perform atlas-based registration.

## Results

Our method (red) converges significantly faster in the first hours of training. Its performance is on par with the best baseline methods. Especially on noisy data, it performs the commonly used supervised and unsupervised $\mathrm{NCC}$ metric.

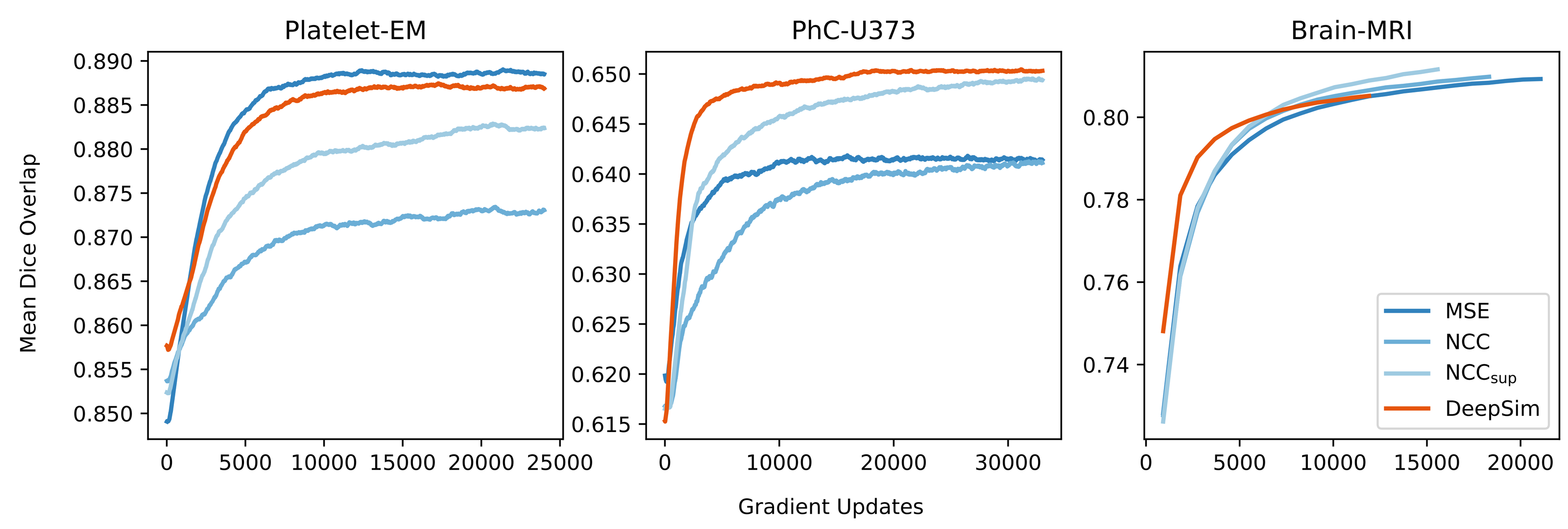


**Figure 2:** Dice Overlap during training on the three datasets.

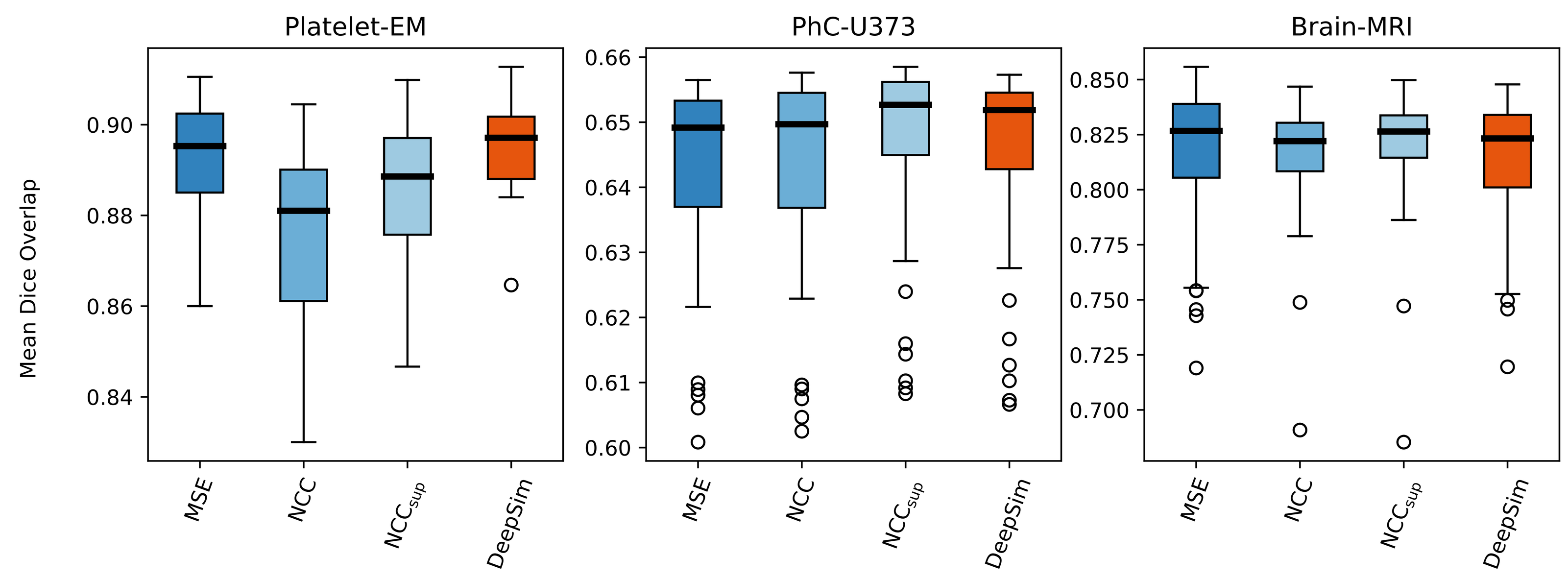


**Figure 3:** Dice overlap on the test sets of the three datasets.

## Ongoing Research

We aim to understand how differences in image anatomy can be detected with the presented metrics. Such an anatomy-aware similarity metric could improve uncertainty estimation and piecewise-diffeomorphic image registration.

## Acknowledgements

This research was supported by the Lundbeck Foundation.

## References


[1] Guha Balakrishnan et. al. VoxelMorph: A Learning Framework for Deformable Medical Image Registration. *IEEE Transactions on Medical Imaging*, 2019.

[2] Richard Zhang et. al. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. *CVPR*, 2018.