Semantic similarity metrics for learned image registration

Steffen Czolbe, Oswin Krause, Aasa Feragen

UNIVERSITY OF COPENHAGEN

DTU

MIDL Lübeck 2021

Semantic Similarity Metric

The semantic similarity metric, titled *DeepSim*, compares semantic representations of the images.

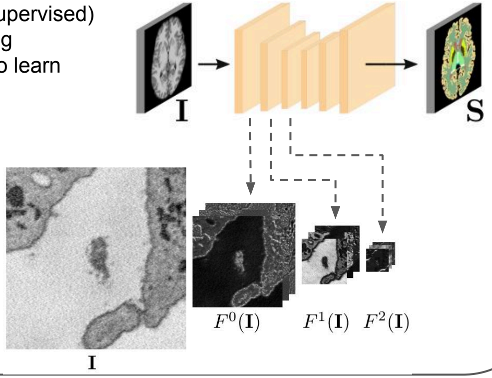
Similarity is calculated pixel-wise with cosine similarity, averaged over the pixels of the images and over multiple layers of abstractions.

Semantic representations are extracted from a network trained on a surrogate task (below).

$$\mathrm{DeepSim}\left(\mathbf{I},\mathbf{J}\right)=\underbrace{\frac{1}{L}\sum_{l=1}^{L}}_{\text{Average over } L \text{ layers of abstraction}}\underbrace{\frac{1}{|\Omega_l|}\sum_{\mathbf{p}\in\Omega_l}}_{\text{Average over }\mathbf{p}\text{ pixels}}\underbrace{\frac{\left\langle F^l_{\mathbf{p}}(\mathbf{I}),F^l_{\mathbf{p}}(\mathbf{J})\right\rangle}{\|F^l_{\mathbf{p}}(\mathbf{I})\|\|F^l_{\mathbf{p}}(\mathbf{J})\|}}_{\text{Cosine similarity of semantic feature representations}}$$

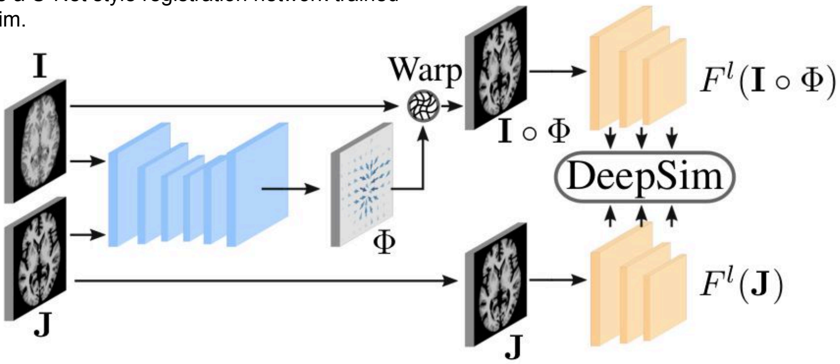
Feature Extraction

We use surrogate tasks like segmentation (supervised) and autoencoding (unsupervised) to learn semantic feature representations.



Registration Model for Evaluation

We evaluate a U-Net style registration network trained with DeepSim.

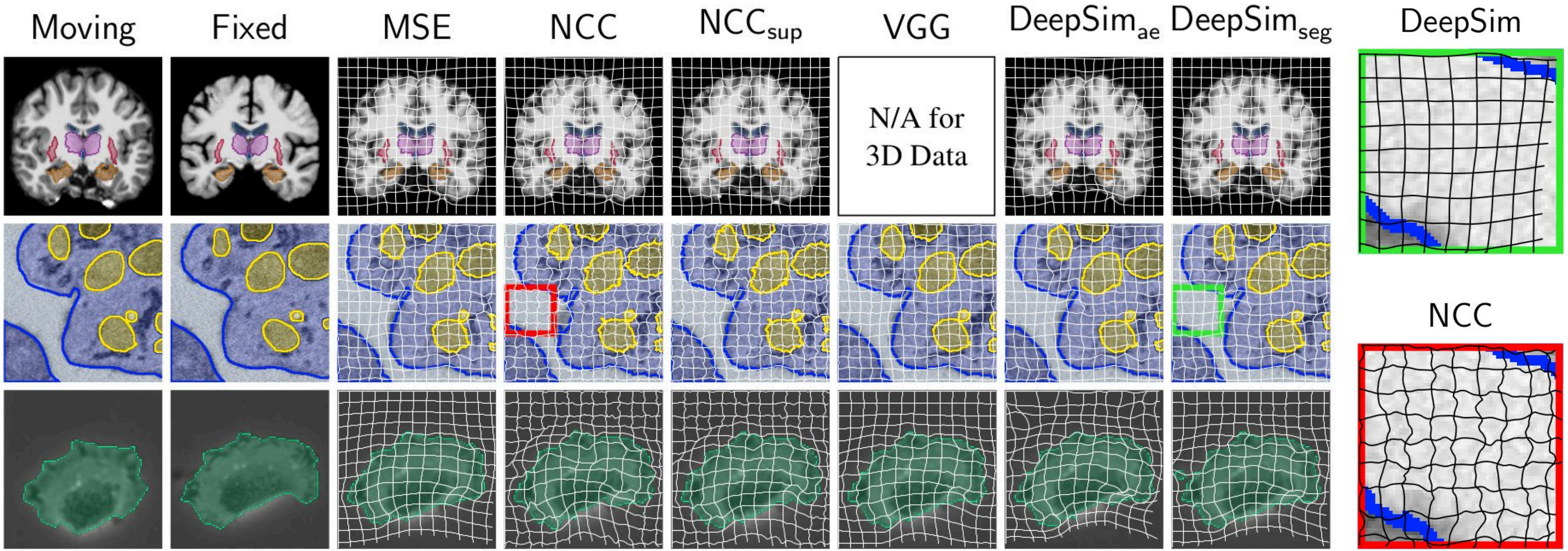


Semantic similarity metrics for learned image registration

Steffen Czolbe, Oswin Krause, Aasa Feragen

Qualitative Comparison

We empirically compare registration models trained with the unsupervised DeepSim$_{ae}$ and semi-supervised DeepSim$_{seg}$ to the baselines MSE, NCC, NCC$_{sup}$(NCC with supervised information), and VGG (a VGG-net based deeps imilarity metric from image generation). Regularization hyperparameters have been tuned for each model separately. We observe smoother transformation fields of DeepSim in areas of high noise (highlighted).



Semantic similarity metrics for learned image registration

Steffen Czolbe, Oswin Krause, Aasa Feragen

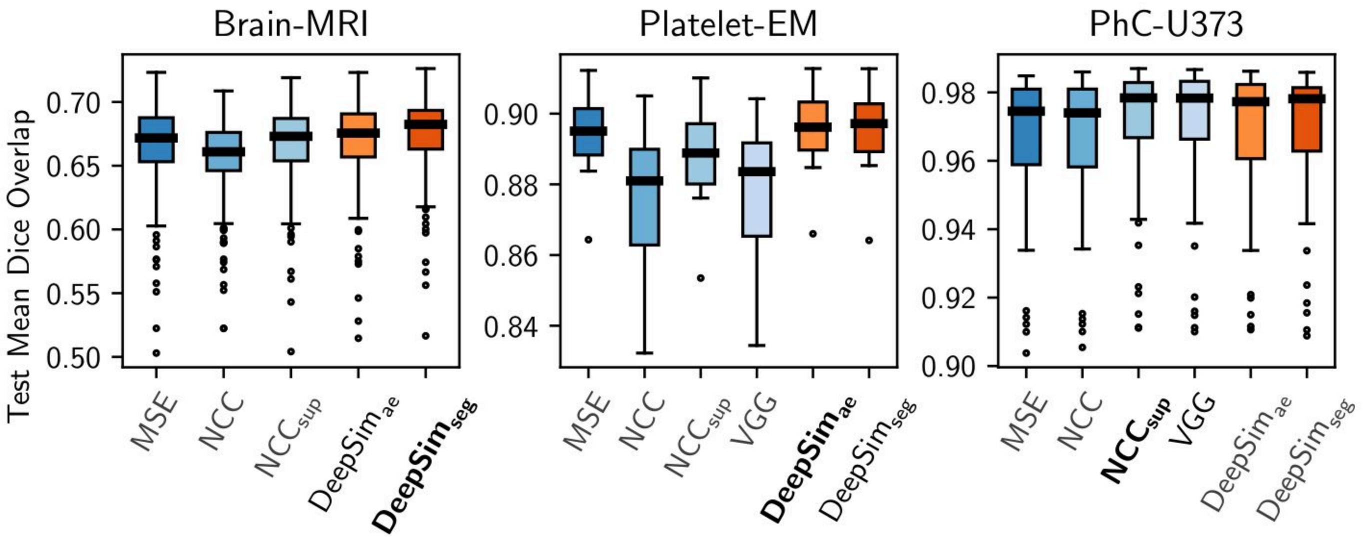
UNIVERSITY OF COPENHAGEN

DTU

MIDL Lübeck 2021

Quantitative Comparison

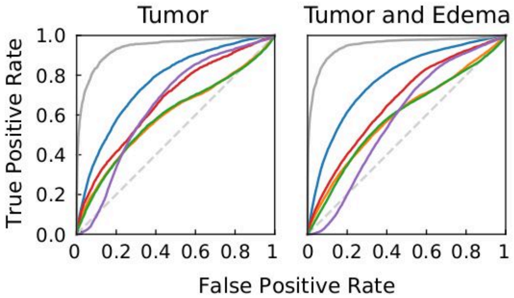
We compare mean segmentation dice overlap on the test set. Baselines in shades of blue, ours in red. Label of the best metric in bold, 2nd best black, others in grey. Boxplot with median, quartiles,deciles and outliers.



Ongoing Work

Semantic similarity metrics improve registration performance in many areas. In a current preprint [1], a registration model trained with the semantic similarity metric outperforms baselines trained with MSE by a large margin.

[1] Steffen Czolbe, Aasa Feragen and Oswin Krause. “Spot the Difference: Topological Anomaly Detection via Geometric Alignment”, 2021. Arxiv Preprint.



Method	AUC (T)	AUC (T+E)
Ours (Sem. Loss)	$.754 \pm .010$	$.774 \pm .003$
Ours (MSE)	$.582 \pm .011$	$.601 \pm .005$
Li and Wyatt	$.672 \pm .012$	$.668 \pm .013$
Jac. Det.	$.590 \pm .009$	$.595 \pm .010$
An and Cho	$.670 \pm .011$	$.584 \pm .013$
Supervised Seg.	$.933 \pm .016$	$.955 \pm .007$