DeepSim: Semantic similarity metrics for learned image registration

Steffen Czolbe **Oswin Krause**
Department of Computer Science
University of Copenhagen
{per.sc, oswin.krause}@di.ku.dk

Aasa Feragen
DTU Compute
Technical University of Denmark
afhar@dtu.dk

Abstract

We propose a semantic similarity metric for image registration. Existing metrics like euclidean distance or normalized cross-correlation focus on aligning intensity values, giving difficulties with low intensity contrast or noise. Our semantic approach learns dataset-specific features that drive the optimization of a learning-based registration model. Comparing to existing unsupervised and supervised methods across multiple image modalities and applications, we achieve consistently high registration accuracy and faster convergence than state of the art, and the learned invariance to noise gives smoother transformations on low-quality images.

1 Introduction

Deformable registration is a fundamental preprocessing tool in medical imaging, where the goal is to find anatomical correspondences between images and derive geometric transformations Φ to align them. Most algorithmic and deep learning-based methods optimize alignment via a similarity measure D and a λ-weighted regularizer R, combined in a loss function

$$L(\mathbf{I}, \mathbf{J}, \Phi) = D(\mathbf{I} \circ \Phi, \mathbf{J}) + \lambda R(\Phi) \ . \tag{1}$$

The similarity metric D assesses the alignment and strongly influences the result. Pixel-based similarity metrics like euclidean distance (MSE) and patch-wise normalized cross-correlation (NCC) are commonly used in both algorithmic [3, 4, 10, 22, 23, 25] and deep learning based [1, 5, 7, 8, 13, 14, 17, 18, 26–28] image registration. Typically, the similarity measure for a task is selected as the best out of a small set of metrics, with no guarantee that one of the metrics is suitable for the data.

The shortcomings of pixel-based similarity metrics have been studied substantially in the image generation community [29], where the introduction of deep similarity metrics approximating human visual perception has improved the generation of photo-realistic images [6, 12]. As registration models are generative models [7], we expect these similarity metrics to improve registration as well. Current attempts at using learned similarity metrics for image registration require ground truth transformations [11] or limit the input to the registration model [17].

We propose a data-driven similarity metric for image registration based on the alignment of semantic features. We learn semantic filters of our metric on the dataset, use it to train a registration model, and validate our approach on three biomedical datasets of different image modalities and applications. Across all datasets, our method achieves consistently high registration accuracy, outperforming even metrics utilizing supervised information. Our models converge faster and learn to ignore noisy image patches, leading to smoother transformations on low-quality data.

Medical Imaging Meets NeurIPS. 34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

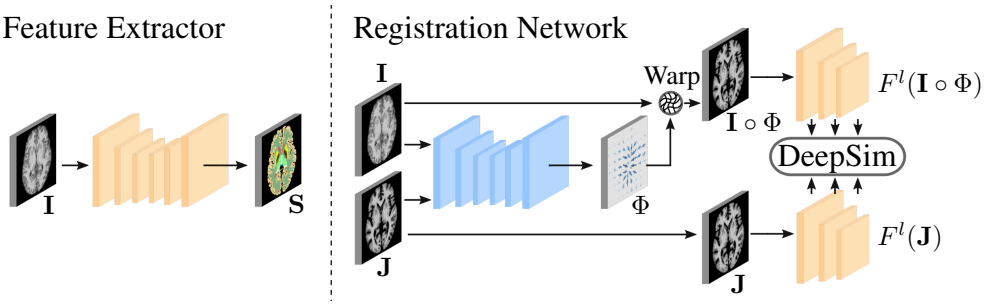

Figure 1: Two-step training: First, the Feature Extractor (yellow) is trained on a segmentation task. Next, its weights are frozen and used in the loss computation of the registration network (blue).

2 A deep similarity metric for image registration

To align areas of similar semantic value, we propose a similarity metric based on the agreement of semantic feature representations of two images. Semantic feature maps are obtained by a *feature extractor* to be tuned on a surrogate segmentation task. To capture alignment of both localized, concrete features, and global, abstract ones, we calculate the similarity at multiple layers of abstraction.

Concretely, 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, we define

$$\mathrm{DeepSim}(\mathbf{I} \circ \Phi, \mathbf{J}) = \frac{1}{L} \sum_{l=1}^{L} \frac{1}{|\Omega_l|} \sum_{\mathbf{p} \in \Omega_l} \frac{\left\langle F^l_{\mathbf{p}}(\mathbf{I} \circ \Phi), F^l_{\mathbf{p}}(\mathbf{J}) \right\rangle}{\|F^l_{\mathbf{p}}(\mathbf{I} \circ \Phi)\| \|F^l_{\mathbf{p}}(\mathbf{J})\|} \quad , \tag{2}$$

where $F^l_{\mathbf{p}}(\mathbf{J})$ denotes the l^{th} layer feature extractor applied to image $\mathbf{J}$, at spatial coordinate $\mathbf{p}$. It is a vector of C_l output channels, and the spatial size of the l^{th} feature map is denoted by $|\Omega_l|$. The metric is influenced by the neighbourhood of a pixel, as F^l composes convolutional filters with increasing receptive area of the composition. Note that the formulation via cosine similarity is similar to the classic NCC metric, which can be interpreted as the squared cosine-similarity between two zero-mean vectors of patch descriptions.

Feature extraction To aid registration, the functions $F^l(\cdot)$ should extract features of semantic relevance for the registration task, while ignoring noise and artifacts. We achieve this by training the feature extractor on a supplementary segmentation task, as segmentation models excel at learning relevant kernels for the data while attaining invariance towards non-predictive features like noise. The obtained convolutional filters act as feature extractors for DeepSim, see also Figure 1.

3 Experiments

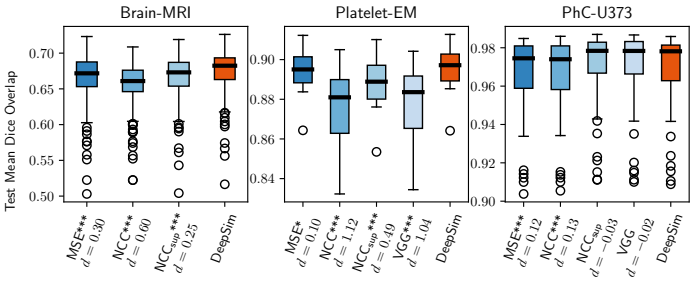

Figure 2: Quantitative comparison of similarity metrics. Stars indicate p-test significance level. Effect size given by Cohen's d.

We compare registration models trained with DeepSim to the baselines MSE, NCC, $\mathrm{NCC}_{\mathrm{sup}}$ (NCC with supervised information [5]), and VGG (a VGG-net based metric common in image generation and similar to our method [12]). Figure 1 shows our model architecture. For both registration and segmentation we use U-nets [21]. The registration network predicts the transformation Φ based on two images $\mathbf{I}, \mathbf{J}$. A spatial transformer module [15] applies Φ to obtain the morphed image $\mathbf{I} \circ \Phi$. The loss is given by Eq. 1; we choose the diffusion regularizer for R and tune hyperparamter λ on the validation sets.

To show that our approach is applicable to a large variety of registration tasks, we validate it on three 2D and 3D datasets of different image modalities: T1-weighted *Brain-MRI* scans [9, 16], human blood cells of the *Platelet-EM* dataset [20], and cell-tracking of the *PhC-U373* dataset [19, 24]. Each dataset is split into a train, validation, and test section.

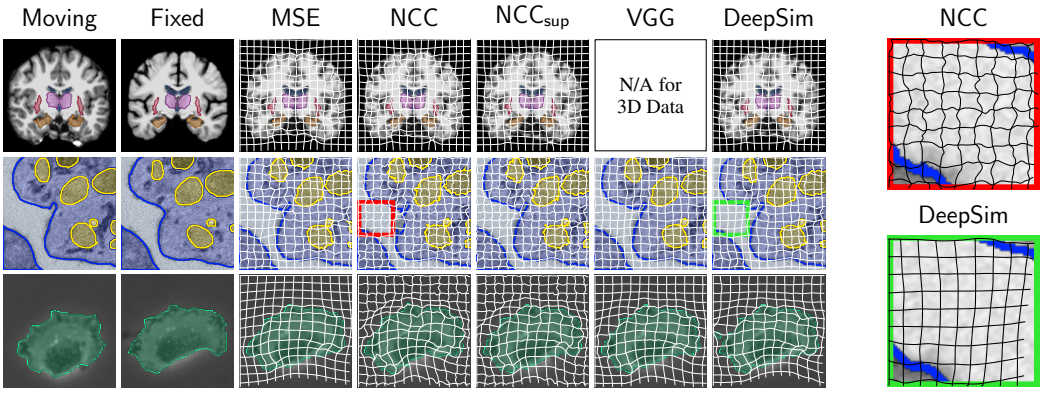

Figure 3: (a) Qualitative comparison, (b) Detail view of highlighted areas. Select segmentation classes annotated color. The transformation is visualized by morphed grid-lines.

Registration accuracy & convergence We measure the mean Sørensen Dice coefficient on the unseen test-set in Figure 2, and test for statistical significance of the result with the Wilcoxon signed rank test for paired samples. Our null hypothesis for each similarity metric is that the model trained with DeepSim performs better. We test for a statistical significance levels of $p^* = 0.05, p^{**} = 0.01, p^{***} = 0.001$. We further measure the effect size with Cohen's d, and label the metrics accordingly in Figure 2. Models trained with our proposed DeepSim rank as the best on the Brain-MRI and Platelet-EM datasets, with strong statistical significance. On the PhC-U373 dataset, all models achieve high dice-overlaps of > 0.97. DeepSim converges faster than the baselines, especially in the first epochs of training.

Qualitative examples & transformation grids We plot the fixed and moving images $\mathbf{I}, \mathbf{J}$ and the morphed image $\mathbf{I} \circ \Phi$ for each similarity metric model in Figure 3a, and a more detailed view of a noisy patch of the Platelet-EM dataset in Figure 3b. The transformation is visualized by grid-lines, which have been transformed from a uniformly spaced grid. On models trained with the baselines, we find strongly distorted transformation fields in noisy areas of the images. In particular, models trained with NCC and NCC_{sup} produce very irregular transformations, despite careful tuning of the regularization-hyper-parameter. The model trained by DeepSim is more invariant towards the noise.

4 Discussion & Conclusion

Registration models trained with DeepSim achieve high registration accuracy across multiple datasets, leading to improved downstream analysis and diagnosis in medical applications. The consistency of our proposed metric makes testing multiple traditional metrics unnecessary; instead of empirically determining whether MSE or NCC captures the characteristics of a data-set best, we can use DeepSim to learn the relevant features from the data.

The analysis of noisy patches in Figure 3b highlights a learned invariance to noise. The pixel-based similarity metrics are distracted by artifacts, leading to overly-detailed transformation fields. DeepSim does not show this problem. While smoother transformation fields can be obtained for all metrics by strengthening the regularizer, this would negatively impact the registration accuracy of anatomically significant regions. Accurate registration of noisy, low-quality images allows for shorter acquisition time and reduced radiation dose in medical applications.

DeepSim is a general metric, applicable to image registration tasks of all modalities and anatomies. Beyond the presented datasets, the good results on low-quality data let us hope that DeepSim will improve registration accuracy in the domains of lung CT and ultrasound, where details are hard to identify, and image quality is often poor. We further emphasize that the application of DeepSim is not limited to deep learning. Algorithmic image registration uses a similar optimization framework, where a similarity-based loss is minimized via gradient descent-based methods. DeepSim can be applied to drive algorithmic methods, improving their performance by aligning deep, semantic feature embeddings.

Acknowledgments and Disclosure of Funding

The authors acknowledges support by the Lundbeck Foundation grant number R218-2016-883.

Broader impact

The broader impact of our work is defined by the numerous applications of medical image registration. Common applications are in neuroscience [5], CT-imaging of lungs and abdomen [26], as well as for fusion and combination of multiple modalities [11].

The deep learning approach to image registration utilized in this work can achieve impressive results across a wide variety of tasks, but this often comes at the cost of training models on specialized hardware for extensive periods. This energy-intensive workload may raise carbon emissions, the primary contributor to climate change [2]. We hope that by presenting a method for learning a semantic similarity metric from the data, we make excessive testing of other loss functions unnecessary. This can reduce the amount of model configurations to be tested in the development of deep learning methods, contributing to a lower environmental impact of the image registration community.

References

[1] Jennifer Alvén, Kerstin Heurling, Ruben Smith, Olof Strandberg, Michael Schöll, Oskar Hansson, and Fredrik Kahl. "A Deep Learning Approach to MR-less Spatial Normalization for Tau PET Images". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019, pp. 355–363.

[2] Lasse F. Wolff Anthony, Benjamin Kanding, and Raghavendra Selvan. "Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models". In: *ICML Workshop on Challenges in Deploying and monitoring Machine Learning Systems* (2020).

[3] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee. "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain". In: *Medical Image Analysis* 12.1 (2008), pp. 26–41.

[4] Brian B. Avants, Nicholas J. Tustison, Gang Song, Philip A. Cook, Arno Klein, and James C. Gee. "A reproducible evaluation of ANTs similarity metric performance in brain image registration". In: *NeuroImage* 54.3 (2011), pp. 2033–2044.

[5] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. "VoxelMorph: A Learning Framework for Deformable Medical Image Registration". In: *IEEE Transactions on Medical Imaging* 38.8 (2019), pp. 1788–1800.

[6] Steffen Czolbe, Oswin Krause, Ingemar Cox, and Christian Igel. "A Loss Function for Generative Neural Networks Based on Watson's Perceptual Model". In: *Advances in Neural Information Processing Systems* (2020).

[7] Adrian V. Dalca, Guha Balakrishnan, John Guttag, and Mert R. Sabuncu. "Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration". In: *Medical Image Computing and Computer Assisted Intervention* (2018), pp. 729–738.

[8] Adrian V. Dalca, Marianne Rakic, John Guttag, and Mert R. Sabuncu. "Learning Conditional Deformable Templates with Convolutional Networks". In: *Advances in neural information processing systems* (2019), pp. 806–818.

[9] Adriana Di Martino, Chao-Gan Yan, Qingyang Li, et al. "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism". In: *Molecular psychiatry* 19.6 (2014), pp. 659–667.

[10] Mirza Faisal Beg, Michael I Miller, Alain Trouvétrouv, and Laurent Younes. "Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms". In: *International Journal of Computer Vision* 61.2 (2005), pp. 139–157.

[11] Grant Haskins, Jochen Kruecker, Uwe Kruger, Sheng Xu, Peter A. Pinto, Brad J. Wood, and Pingkun Yan. "Learning deep similarity metric for 3D MR–TRUS image registration". In: *International Journal of Computer Assisted Radiology and Surgery* 14.3 (2019), pp. 417–425.

[12] Xianxu Hou, Linlin Shen, Ke Sun, and Guoping Qiu. "Deep feature consistent variational autoencoder". In: *Winter Conference on Applications of Computer Vision*. IEEE. 2017, pp. 1133–1141.

[13] Xiaojun Hu, Miao Kang, Weilin Huang, Matthew R. Scott, Roland Wiest, and Mauricio Reyes. "Dual-Stream Pyramid Registration Network". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019, pp. 382–390.

[14] Yipeng Hu, Eli Gibson, Dean C. Barratt, Mark Emberton, J. Alison Noble, and Tom Vercauteren. "Conditional Segmentation in Lieu of Image Registration". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019, pp. 401–409.

[15] Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. "Spatial Transformer Networks". In: *Advances in neural information processing systems* (2015).

[16] Pamela J LaMontagne, Tammie L S Benzinger, John C Morris, et al. "OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease". In: *medRxiv* (2019).

[17] Matthew C. H. Lee, Ozan Oktay, Andreas Schuh, Michiel Schaap, and Ben Glocker. "Image-and-Spatial Transformer Networks for Structure-Guided Image Registration". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019, pp. 337–345.

[18] Lihao Liu, Xiaowei Hu, Lei Zhu, and Pheng-Ann Heng. "Probabilistic Multilayer Regularization Network for Unsupervised 3D Brain Image Registration". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019, pp. 346–354.

[19] Martin Maška et al. "A benchmark for comparison of cell tracking algorithms". In: *Bioinformatics* 30.11 (2014), pp. 1609–1617.

[20] Matthew Quay, Zeyad Emam, Adam Anderson, and Richard Leapman. "Designing deep neural networks to automate segmentation for serial block-face electron microscopy". In: *International Symposium on Biomedical Imaging*. Vol. 2018-April. IEEE Computer Society, 2018, pp. 405–408.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Vol. 9351. Springer Verlag, 2015, pp. 234–241.

[22] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. "Nonrigid registration using free-form deformations: Application to breast mr images". In: *IEEE Transactions on Medical Imaging* 18.8 (1999), pp. 712–721.

[23] Jean-Philippe Thirion. *Image matching as a diffusion process: an analogy with Maxwell's demons*. Tech. rep. 3. 1998, pp. 243–260.

[24] Vladimír Ulman et al. "An objective comparison of cell-tracking algorithms". In: *Nature Methods* 14.12 (2017), pp. 1141–1152.

[25] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. "Non-parametric diffeomorphic image registration with the demons algorithm". In: *Lecture Notes in Computer Science*. Vol. 4792 LNCS. PART 2. 2007, pp. 319–326.

[26] Bob D. de Vos, Floris F. Berendsen, Max A. Viergever, Hessam Sokooti, Marius Staring, and Ivana Išgum. "A deep learning framework for unsupervised affine and deformable image registration". In: *Medical Image Analysis* 52 (2019), pp. 128–143.

[27] Zhenlin Xu and Marc Niethammer. "DeepAtlas: Joint Semi-supervised Learning of Image Registration and Segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. 2019, pp. 420–429.

[28] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. "Quicksilver: Fast predictive image registration - A deep learning approach". In: *NeuroImage* 158 (2017), pp. 378–396.

[29] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric". In: *Conference on Computer Vision and Pattern Recognition* (2018), pp. 586–595.