Continuous benchmarking in medical image registration - review of the current state of the Learn2Reg challenge

Lasse Hansen*[1] hansen@imi.uni-luebeck.de
[1] *Institute of Medical Informatics, Universität zu Lübeck, Lübeck, Germany*
Alessa Hering*[23] alessa.hering@mevis.fraunhofer.de
[2] *Fraunhofer MEVIS, Bremen, Germany*
[3] *Diagnostic Image Analysis Group, Radboud University Medical Center, Nijmegen, Netherlands*
Christoph Großbröhmer[1] grossbroehmer@imi.uni-luebeck.de
Mattias P. Heinrich[1] heinrich@imi.uni-luebeck.de

Editors: Under Review for MIDL 2022

Abstract

Image registration is a fundamental medical image analysis task, and a wide variety of approaches have been proposed. However, only a few studies have comprehensively compared medical image registration approaches on a wide range of clinically relevant tasks, in part because of the lack of availability of such diverse data. This limits the development of registration methods, the adoption of research advances into practice, and a fair benchmark across competing approaches. The Learn2Reg challenge addresses these limitations by providing a multi-task medical image registration benchmark for comprehensive characterisation of deformable registration algorithms. We established an easily accessible framework for training and validation of 3D registration methods, which so far enabled the compilation of results of over 65 individual method submissions from more than 20 unique teams. We used a complementary set of metrics, including robustness, accuracy, plausibility, and runtime, enabling unique insight into the current state-of-the-art of medical image registration. In this abstract for the MIDL community we want to 1) give a shortest (graphical) overview of the Learn2Reg Challenge, 2) present key results and outcomes of past editions and 3) outline limitations and resulting ongoing work.[†]

1. Learn2Reg challenge overview

Currently, Learn2Reg comprises six clinically relevant, complementary tasks (datasets) covering a wide range of anatomies (brain, abdomen and thorax), modalities (ultrasound, CT, MR), availability of annotations, as well as intra- and inter-patient registration evaluation (summarised in Table 1). New tasks are introduced as part of the annual challenge workshops and are available from that point on for continuous algorithm development at learn2reg.grand-challenge.org (general information, publicly available training data and automatic evaluation of displacement fields on validation splits) and learn2reg-test.grand-challenge.org (evaluation of dockerised algorithms on hidden test sets).

[*] Contributed equally
[†] Full paper, author list and affiliations: (Hering et al. (2021), arxiv.org/abs/2112.04489)

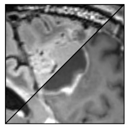
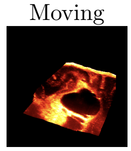
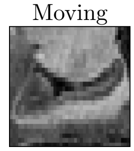
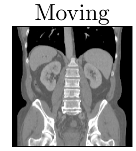
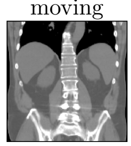
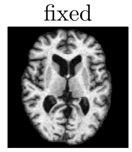
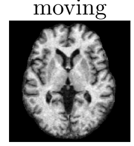
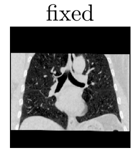
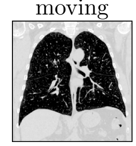
	CuRIOUS		Hippocampus MR		Abdomen CT-CT	
	Fixed	Moving	Fixed	Moving	Fixed	Moving
						
Modalities	MR T1w & FLAIR/US		MR T1w/MR T1w		CT/CT	
Intra-/Inter-patient	Intra-patient		Inter-patient		Inter-patient	
Cases (Train/Test)	32 (22/10)		394 (263/131)		50 (30/20)	
Challenges	● ● ●		●		● ●	
	Abdomen MR-CT		OASIS		Lung CT	
	fixed	moving	fixed	moving	fixed	moving
						
Modalities	MR T1w / CT		MR T1w / MR T1w		CT / CT	
Intra-/Inter-patient	Intra-patient		Inter-patient		Intra-patient	
Cases (Train/Test)	16 (8/8)		455 (416/39)		30 (20/10)	
Challenges	● ● ● ●		●		● ● ● ●	

Table 1: Overview of the current six Learn2Reg tasks adressing imminent challenges of medical image registration: multi-modal scans ●, few/noisy annotations ●, partial visibility ●, small datasets ●, large deformations ●, small structures ●, unsupervised registration ● and missing correspondences ●.

2. Key results and outcomes

For the comprehensive evaluation of submitted methods to the first two editions of Learn2Reg (2020 and 2021), we considered a number of complementary metrics that assess the accuracy (Dice similarity coefficient (DSC), Hausdorff distance (HD), target registration error (TRE)), robustness (percentiles of accuracy metrics), plausibility (standard deviation of the log Jacobian determinant), and runtime of the algorithms, which enabled unique insights into the current state-of-the-art for medical image registration.

Successfull benchmarking The Learn2Reg challenge attracted 20 international teams with a total of 65 individual methods submitted for the six tasks. This represents one of the most comprehensive comparisons of registration algorithms and lays the foundation for a fair standardised benchmark in medical image registration.

DL vs. conventional registration An interesting result of the challenge is that many top-performing methods, including the 3 overall winners (Siebert et al., 2022; Mok and Chung, 2020; Häger et al., 2021), have chosen a combination of machine learning based and conventional registration (both for different tasks and within single tasks). Thus, in contrast to other areas of medical image processing, such as semantic segmentation, no definite dominance of Deep Learning is evident. A clear trend is the use of DL-based initial registration followed by conventional (GPU-supported) instance optimisation.

3. Limitations and ongoing work

The first two versions of Learn2Reg have provided important steps towards a fair comparison of registration algorithms, nevertheless, the challenge still has some limitations. Firstly, registration accuracy cannot be measured directly but must be evaluated via auxiliary metrics such as the overlap of segmentation masks. While this is an inherent problem in evaluating registration, this issue can be mitigated by generating further manual annotations for certain structures. Second, the amount of available annotated training data varied across tasks and made in particular intra-patient tasks harder for learning-based approaches. Unfortunately, the problem is that large datasets are often not publicly available and therefore cannot be used in this Type 1 challenge. Third, for most tasks, all segmentation classes used for testing were also available in the training. This was due to the fact, that for three out of four tasks with segmentation labels these annotations were already publicly available prior to Learn2Reg and we considered it in-transparent (and biased) to simply not point participants to their availability. And finally, statements about the quality of the registration algorithms can only be generalised to a limited extent, but apply mainly to the selected tasks.

Availability/Continuity We have set up a continuous evaluation on the test data of all tasks on learn2reg-test.grand-challenge.org. This enables the continuity of the challenge and increase the transparency and availability of algorithms since the algorithms have to be uploaded as docker containers to grand-challenge.

Generalisability Now that steps towards comprehensive registration benchmarks are made, we consider finding a self-configuring registration framework similar to the nnU-Net framework the next important objective. Learn2Reg 2022 sets out to become the first Type 2 medical registration challenge with hidden training data and the requirement to submit training algorithms that auto-tune hyperparameters for unseen tasks.

References

S Häger, S Heldmann, A Hering, A Kuckertz, and A Lange. Variable fraunhofer MEVIS RegLib comprehensively applied to Learn2Reg challenge. In *MICCAI Workshops*, 2021.

A Hering, L Hansen, T Mok, A Chung, H Siebert, S Häger, A Lange, S Kuckertz, S Heldmann, W Shao, S Vesal, M Rusu, G Sonn, T Estienne, M Vakalopoulou, L Han, Y Huang, M Brudfors, Y Balbastre, S Joutard, M Modat, G Lifshitz, D Raviv, J Lv, Q Li, V Jaouen, D Visvikis, C Fourcade, M Rubeaux, W Pan, Z Xu, B Jian, F De Benetti, M Wodzinski, N Gunnarsson, J Sjölund, H Qiu, Z Li, C Großbröhmer, A Hoopes, I Reinertsen, Y Xiao, B Landman, Y Huo, K Murphy, N Lessmann, B van Ginneken, A Dalca, and M Heinrich. Learn2Reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning, 2021. URL `https://arxiv.org/abs/2112.04489`.

T Mok and A Chung. Large deformation diffeomorphic image registration with laplacian pyramid networks. In *MICCAI*, 2020.

H Siebert, L Hansen, and M Heinrich. Fast 3D registration with accurate optimisation and little learning for Learn2Reg 2021. In *MICCAI Workshops*, 2022.