Discrete Unsupervised 3D Registration Methods for the Learn2Reg Challenge

Lasse Hansen and Mattias P. Heinrich

Institute of Medical Informatics, Universität zu Lübeck, Lübeck, Germany
\{hansen,heinrich\}@imi.uni-luebeck.de

Abstract. The Learn2Reg challenge poses four very different tasks with varying difficulty for image registration algorithms. In this short paper, we describe our choices for two state-of-the-art discrete 3D registration methods that enable fast and accurate estimation of large deformations without expert supervision during training. Both approaches primarily focus on the use of contrast-invariant features with dense displacement evaluation and were ranked among the top three of all challenge contestants, yielding two first places and three second places for the four sub-tasks.

Keywords: Discrete optimisation · Graphical models · Contrast-independent features

1 Motivation and Background

Deformable image registration, in particular deep learning based approaches, have often struggled with capturing large deformations of local anatomy. Conventional registration methods either rely on multiple warps and coarse-to-fine resolution schemes, which are hard to mimic within learning based framework due to memory constraints, or discrete optimisation to avoid local minima. Here, we analyse our recently proposed probabilistic dense displacement net (PDD-net) [4,5] in comparison to the state-of-the-art 3D discrete registration framework deeds [7] using modality-invariant self-similarity descriptors [6,9]. Our results demonstrate that despite using no labels for supervision both methods are highly competitive with respect to the best performing conventional and learning-based approaches across all four tasks - with a particular edge for datasets with limited annotations.

2 Methods

In the subsequent two section, we present a brief overview of the methodological principals and algorithmic building blocks of deeds and PDD-net with a focus on their common and distinct features.
2.1 Dense Displacement Sampling (Deeds) with Discrete Spanning Tree Optimisation

The main idea behind deeds is to explore a very large number of degrees of freedom (of the nonlinear deformation) that correspond to potential discretised displacements. In order to obtain a tractable computational complexity the following three approximations to the exact discrete optimisation problem are made. 1) instead of directly solving the deformation estimation with a single quantised warp, several levels of decreasingly coarse control point grids are employed, 2) patch-based similarity metrics are approximated using a quantised range of values (and the efficient Hamming distance) and a subsampling of voxels, 3) a simplified graph-model, namely a minimum-spanning-tree is employed (that only requires a single forward and backward path of messages for optimisation) in combination with a symmetric inverse consistency approach (see details in [7,9]). By employing the robust yet accurate hand-crafted MIND self-similarity context (SSC) descriptors, the method is applicable to multimodal tasks as well as challenging image appearance, e.g. due to respiration in lung CT or varying contrast in abdominal CT. deeds and its extensions (e.g. keypoint matching in [11]) have excelled in lung registration (first place at EMPIRE10 and for the DIR-lab COPD dataset), abdominal registration (MICCAI 2015 beyond the cranial vault challenge) and MRI-US fusion (CuRIOUS challenge MICCAI 2018,2019).

2.2 Probabilistic Dense Displacement (PDD) Net

The PDD-net aims to mimic the successful discrete components of deeds, while reducing the run-time through GPU implementation by over an order of magnitude and enabling unsupervised feature learning. We firstly use a compact deformable convolutional network [8] to extract features and compute a dense 6D (3 spatial + 3 displacement dimensions) dissimilarity tensor. Employing the minimum-spanning-tree optimisation of deeds within a deep learning framework would require the back propagation through dozens of sequential message passing steps. We therefore opted for a simpler graphical model, the mean field inference for a conditional random field (different to Markov random fields no directed messages have to be computed), which enables message passing by (Gaussian) filtering operations over the three spatial dimensions. The label compatibility function (see [2,10] for details) is defined as approximated min-convolutions: a combination of min-filtering and smoothing that act on the three displacement dimensions. These steps are interleaved and repeated twice to predict a discrete probabilistic (softmax) displacement map for each control point. As described in [5] these probabilities can be effectively employed for unsupervised (metric) learning using the aforementioned MIND-SSC descriptors. In most scenarios a single warp of this approach yields very accurate results in less than a second that can be further finetuned using continuous instance optimisation or a second warp.
| Task       | Task1 Score | Task1 Rank | Task2 Score | Task2 Rank | Task3 Score | Task3 Rank | Task4 Score | Task4 Rank | Overall Score | Overall Rank |
|------------|-------------|------------|-------------|------------|-------------|------------|-------------|------------|---------------|--------------|
| LapIRN     | 0.72        | 3          | 0.75        | 3          | 0.93        | 1          | 0.95        | 1          | 0.83          | 1            |
| PDD        | 0.94        | 1          | 0.85        | 1          | 0.69        | 4          | 0.74        | 4          | 0.80          | 2            |
| Deeds      | 0.91        | 2          | 0.79        | 2          | 0.73        | 2          | 0.55        | 6          | 0.73          | 3            |
| LibReg     | 0.48        | 4          | 0.71        | 5          | 0.50        | 7          | 0.64        | 5          | 0.57          | 4            |
| Uppsala    | 0.45        | 5          | 0.54        | 6          | 0.60        | 6          | 0.45        | 7          | 0.51          | 5            |
| CentraleSupélec | 0.39 | 6          | 0.25        | 8          | 0.73        | 2          | 0.77        | 3          | 0.49          | 6            |
| Nifty      | 0.39        | 6          | 0.72        | 4          | 0.35        | 8          | 0.43        | 8          | 0.45          | 7            |
| AGH        | 0.39        | 6          | 0.25        | 8          | 0.21        | 9          | 0.78        | 2          | 0.36          | 8            |
| KCL        | 0.39        | 6          | 0.25        | 8          | 0.68        | 5          | 0.20        | 9          | 0.34          | 9            |
| EC Nantes  | 0.39        | 6          | 0.51        | 7          | 0.21        | 9          | 0.20        | 9          | 0.30          | 10           |

3 Experiments and Results

For each of the four distinct Learn2Reg tasks we explain the experimental choices and important hyperparameters. Subsequently, we report numerical results for each task. Quantitative challenge results can be found in Table 1. The table shows scores and ranks for each task as well as for the entire challenge. The scores represent a combined normalized evaluation criterion that includes metrics such as Dice similarity, target registration error (TRE), smoothness of the transformation (standard deviation of log Jacobian) and runtime of the algorithm. Final ranks are based on significance of scores. Further numerical results can be found at the Learn2Reg website\(^1\). Figure 1 shows examples of registration results of our contributions and selected comparison methods for each task.

*Task 1 CuRIOUS US/MRI.* This task aims to correct brain shift in multimodal US/MRI images. **deeds** is used with standard settings for affine transformation as described in the publicly available code repository\(^2\). For the **PDD-net** we chose to use handcrafted MIND features, which showed convincing results for the multi-modal image modalities in this task. The task poses the problem of finding an affine transformation between the US and the MRI image. We therefore, estimate an affine matrix from predicted displacements using a trimmed least square approach. For a robust registration it was also necessary to mask the ultrasound image. The registration accuracy is evaluated using manual annotated landmarks. For deeds and the PDD-net the TRE is 3.89 mm and 3.09 mm, respectively. With 4.61 s the PDD-net is almost twice as fast as deeds (9.12 s).

*Task 2 Lung CT.* The second task deals with aligning inhale and exhale lung CT scans. We use default settings (See footnote 2) for the **deeds** framework

---

1. [https://learn2reg.grand-challenge.org](https://learn2reg.grand-challenge.org).
2. [https://github.com/mattiaspaule/deedsBCV](https://github.com/mattiaspaule/deedsBCV).
and additionally mask the lung regions. As in the first task we use fixed MIND features for the PDD-Net. To exploit the importance of lung vessels for this registration task we implement a sparse variant, where instead of relying on a fixed grid for similarity computations, sparse keypoints are extracted from the inhale scan using the Foerstner operator [3]. To deal with the sparse keypoints within the PDD framework the Gaussian filtering on the fixed grid is replaced by Laplacian smoothing on the kNN graph ($k = 10$). Experiments showed that employing a second warp (after transformation of the moving image) yields significant better results. We therefore employ two warps with the PDD-net using 1024 keypoints in the first and 1536 keypoints in the second warp. As in the first task the registration accuracy is evaluated on manual annotated corresponding landmarks in the inhale and exhale scan. With TREs of 2.26 mm and 2.46 mm the results of deeds and the PDD-net are comparable. However, the runtime of the PDD-net is much faster (2.49 s vs. 41.32 s). The best registration accuracy is achieved by the contribution of Fraunhofer MEVIS with 1.72 mm.

**Fig. 1.** Qualitative results of contributions and selected comparison methods for all four tasks of the L2R Challenge.
**Task 3 Abdominal CT.** This task aims at the inter-patient registration of abdominal CT scans. Default settings (see Footnote 2) for deeds showed convincing results. For the PDD registration framework an Obelisk network [8] was used to extract features from the fixed and moving image. The network was trained end-to-end using an unsupervised non-local MIND loss. We employed a second warp after the first transformation of the moving image (using the same feature extraction network). A large boost in registration accuracy with only a small decrease in runtime is observed when using instance optimisation (100 iterations using Adam optimizer) after the initial prediction of displacements with the PDD-net. The registration accuracy is assessed by the Dice similarity of organ segmentation labels. deeds and PDD-net achieve Dice scores of 0.46 and 0.51, while the runtimes are 41.65 s and 4.05 s, respectively. The best registration accuracy of all contestants is achieved by the LapIRN framework with a Dice score of 0.67 (using label supervision).

**Task 4 Hippocampus MRI.** The final task is to align the hippocampus head and body for inter-patient comparison. As the hippocampus MRI images are relatively small (64 × 64 × 64) we needed to adjust the default parameters (see Footnote 2) for deeds. The grid point spacing for the five resolution levels are set to 6, 5, 4, 3 and 2 respectively. For this task the PDD-net is extended by a Voxelmorph framework [1]. The combination of both frameworks gave the best registration results in our experiments. For the first warp, the PDD-net with fixed MIND features is used to predict larger deformations. Then, to cope with remaining small scale transformations, a Voxelmorph network is trained using an unsupervised MIND loss. As in the third task the registration accuracy is evaluated using Dice similarity of segmentation labels (hippocampus head and body). For deeds and the PDD-net the Dice score is 0.76 and 0.78, respectively. The runtime for the PDD-net is 10x faster than for deeds (0.31 s vs. 3.14 s). LapIRN achieves the highest Dice score (0.86) using label supervision.

**4 Conclusion**

In our contributions to the Learn2Reg challenge we analysed the use of two discrete registration methods (deeds, PDD-net) with several experimental design choices (MIND features, Obelisk features, instance optimisation, etc.) for the four distinctive challenge tasks. Both proposed methods were ranked among the top three of all challenge contestants, which establishes them as general registration frameworks. The PDD-net stands out with fast runtimes and winning task 1 and 2 of the challenge, while deeds achieved very consistent registration scores (ranked second for Task 1, 2 and 3).

**References**

1. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imaging 38(8), 1788–1800 (2019)
2. Felzenszwalb, P.F., Huttenlocher, D.P.: Distance transforms of sampled functions. Theory Comput. **8**(1), 415–428 (2012)
3. Förstner, W., Gülch, E.: A fast operator for detection and precise location of distinct points, corners and centres of circular features. In: Intercommission Conference on Fast Processing of Photogrammetric Data, pp. 281–305 (1987)
4. Heinrich, M.P.: Closing the gap between deep and conventional image registration using probabilistic dense displacement networks. In: Shen, D., et al. (eds.) MICCAI 2019. LNCS, vol. 11769, pp. 50–58. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32226-7_6
5. Heinrich, M.P., Hansen, L.: Highly accurate and memory efficient unsupervised learning-based discrete CT registration using 2.5D displacement search. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12263, pp. 190–200. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_19
6. Heinrich, M.P., et al.: Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration. Med. Image Anal. **16**(7), 1423–1435 (2012)
7. Heinrich, M.P., Jenkinson, M., Brady, S.M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Trans. Med. Imaging **32**(7), 1239–48 (2013)
8. Heinrich, M.P., Oktay, O., Bouteldja, N.: Obelisk-net: fewer layers to solve 3D multi-organ segmentation with sparse deformable convolutions. Med. Image Anal. **54**, 1–9 (2019)
9. Heinrich, M.P., Jenkinson, M., Papież, B.W., Brady, S.M., Schnabel, J.A.: Towards realtime multimodal fusion for image-guided interventions using self-similarities. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8149, pp. 187–194. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40811-3_24
10. Krähenbühl, P., Koltun, V.: Efficient inference in fully connected CRFs with Gaussian edge potentials. Adv. Neural Inf. Process. Syst. **24**, 109–117 (2011)
11. Rühaak, J., et al.: Estimation of large motion in lung CT by integrating regularized keypoint correspondences into dense deformable registration. IEEE Trans. Med. Imaging **36**(8), 1746–1757 (2017)