Large Deformation Image Registration with Anatomy-Aware Laplacian Pyramid Networks

Tony C. W. Mok(✉) and Albert C. S. Chung

Lo Kwee-Seong Medical Image Analysis Laboratory,
Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology,
Clear Water Bay, Hong Kong
cwmokab@connect.ust.hk

Abstract. Deep learning-based methods have recently demonstrated remarkable results in deformable image registration for a wide range of medical image analysis tasks. However, most of the deep learning-based approaches are often limited to small deformation settings. In this paper, we describe a deformable image registration approach for the Learn2Reg 2020 challenge based on the Laplacian pyramid image registration networks. Our approach won 1st place in the Learn2Reg 2020 challenge.

Keywords: Image registration · Diffeomorphic registration · Deep Laplacian pyramid networks · Learn2Reg

1 Introduction

Medical image registration is important in a variety of medical image analysis and has been a topic of active research for decades. Recently, several unsupervised deep learning-based approaches [2,3,11,16,19] have been proposed for deformable image registration and achieved remarkable performance in terms of registration accuracy, computation speed and robustness. To evaluate state-of-the-art methods for image registration, the Learn2Reg 2020 challenge [1] consists of four clinically relevant sub-tasks, including brain intra-operative ultrasound (iUS) to MRI registration [17], exhale-to-inhale lung CT registration [9], abdominal CT registration [18], and Hippocampus registration [15]. The clinical subtasks in the Learn2Reg challenge impose three critical challenges: estimating large deformations, learning from small datasets, and dealing with multi-modal scans, which are challenging for deep learning-based approaches.

In this paper, we present an image registration method based on a specific deep convolutional neural network (CNN) architecture for large deformation registration, which won the Learn2Reg 2020 challenge. We adopt the Laplacian pyramid network with anatomical label supervision to overcome the large inter- or intra-variations of the anatomical structures in the input scans, and data augmentation to alleviate the overfitting issue.

N. Shusharina et al. (Eds.): ABCs 2020/L2R 2020/TN-SCUI 2020, LNCS 12587, pp. 61–67, 2021.
https://doi.org/10.1007/978-3-030-71827-5_7

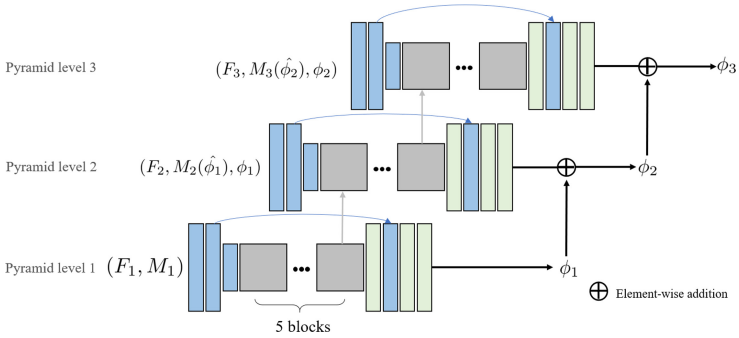

Fig. 1. Overview of the proposed 3-level deep Laplacian pyramid image registration networks. The feature maps from feature encoder, a set of 5 residual blocks, and feature decoder are colored with blue, gray and green, respectively. We highlight that all registrations are done in 3D throughout this paper. For clarity and simplicity, we depict the 2D formulation of our method in the figure. (Color figure online)

2 Methods

Existing deep learning-based approaches [2,3,14,16] often rely on the affine pre-registration and are limited to small deformation settings. Consequently, the registration accuracy degrades whenever there exists large inter- or intra-variation of the anatomical structures in the input scans. Motivated by the successes of deep Laplacian pyramid networks in a variety of computer vision tasks [4,10], we proposed the Laplacian pyramid image registration network (LapIRN) [12] that aims to address the medical image registration problem in large deformation settings.

2.1 Laplacian Pyramid Image Registration Networks

Let F and M denote a fixed 3D scan and a moving 3D scan, respectively. We formulate the image registration problem as a learning problem based on deep CNN. The goal of our method is to estimate the optimal displacement fields ϕ^* such that the dissimilarity between the warped moving scan $M(\phi^*)$ and F are minimized, subject to the smoothness regularization on ϕ^*.

We implement our image registration approach by using a 3-level Laplacian pyramid framework to naturally mimic the conventional multi-resolution strategy. Figure 1 depicts the network architecture of our method. We first create the input image pyramid by downsampling the input images with trilinear interpolation to obtain $F_i \in \{F_1, F_2, F_3\}$ (and $M_i \in \{M_1, M_2, M_3\}$), where F_i denotes the downsampled F with a scale factor $0.5^{(L-i)}$ such that F_1 has the coarsest spatial resolution in F_i and $F_3 = F$. For each level $i \in 1, 2, 3$, we utilize the identical CNN-based registration network (CRN) presented in [12] to capture hierarchical features of the input image pair in pyramid level i (i.e., the concatenation of F_i and M_i) necessary to estimate the correspondence deformation

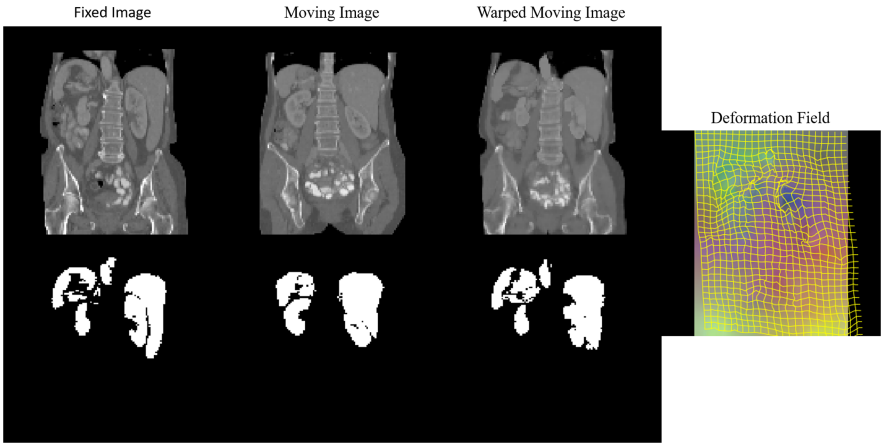

Fig. 2. Example coronal CT slices of inter-patient abdominal CT registration using LapIRN. The corresponding segmentation label maps of the anatomical structures are represented in the second row. Note that we visualize all distinct anatomical labels with the same color for simplicity. (Color figure online)

field ϕ_i at level i. Similar to the conventional multi-resolution optimization approach, the registration starts with the input image pair that has the coarsest spatial resolution (i.e., F_1 and M_1). The resulting deformation field ϕ_1 and the high-level embedding from the CRN in pyramid level i is then passed to the next level. The skip connection between the CRNs at different pyramid levels greatly increases the receptive field as well as the non-linearity of the network to learn complex non-linear correspondence at the finer levels. For pyramid level $i > 1$, the CRN focuses on the misalignment of F_i and $M_i(\hat{\phi}_{i-1})$,and takes the concatenation of F_i, M_i and the upsampled ϕ_{i-1} as input. The output deformation ϕ_i is formed by the addition of the output of CRN in the pyramid level i and added with upsampled deformation field from the previous level $\hat{\phi}_{i-1}$. We repeat this registration process until the finest level completed.

2.2 Anatomical Label Supervision

To overcome the registration difficulties caused by the large inter- or intra-variations of the anatomical structures in the input scans, we introduce the anatomical label supervision along with the similarity pyramid to the loss function of LapIRN. The complete loss is therefore:

$$\mathcal{L}_p = \mathcal{S}^p + \mathcal{L}_{dice} + \frac{\lambda}{2^{(L-p)}} ||\nabla \phi_p||_2^2, \tag{1}$$

where $p \in \{1, 2, 3\}$ denotes the current pyramid level, $\mathcal{S}^p$ represents the similarity pyramid with p levels, $\mathcal{L}_{dice}$ is the Dice score loss of the anatomical segmentation labels, the last term is the smoothness regularization on the displacement fields

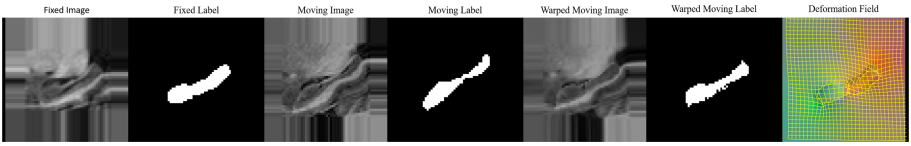

Fig. 3. Example sagittal MR slices of Hippocampus registration using LapIRN.

ϕ_p, and λ is a regularization parameter. The similarity pyramid is a weighted sum of the similarity between the fixed image and warped moving image in different spatial resolutions, which is less sensitive to the image noise and helps to avoid local minimal solutions. Mathematically, The proposed similarity pyramid is formulated as:

$$\mathcal{S}^K(F, M) = \sum_{i \in [1..K]} -\frac{1}{2^{(K-i)}} NCC_w(F_i, M_i), \tag{2}$$

where $\mathcal{S}^K(\cdot, \cdot)$ denotes the similarity pyramid with K levels, NCC_w represents the local normalized cross-correlation with windows size w^3, and (F_i, M_i) denotes the images in the image pyramid. Although the similarity pyramid is robust to the noise, it heavily relies on the intensity-based similarity metric, which implies that this metric cannot differentiate distinct anatomical structures with similar intensity values presented in the input scans. As such, we decide to utilize the anatomical segmentation label in the dataset and introduce an anatomy-aware loss term $\mathcal{L}_{dice}$. Given the anatomical segmentation labels of the fixed image F_i^{seg} and the moving image M_i^{seg} in i-th pyramid level. The proposed anatomy-aware loss term is formulated as:

$$\mathcal{L}_{dice}(F_i^{seg}, M_i^{seg}(\phi_i)) = \frac{2|F_i^{seg} \cap M_i^{seg}(\phi_i)|}{|F_i^{seg}| + |M_i^{seg}(\phi_i)|}. \tag{3}$$

2.3 Data Preprocessing and Augmentation

We adopt a simple preprocessing pipeline for the image scans in all sub-tasks. We downsample all the image scans with a factor of 2 and normalized the intensity values to $[0, 1]$. For task 3, we also apply the windowing with lower and upper bound set to -500 and 800 respectively on the abdominal CT scans. During the training phase of our method, we apply affine augmentation with a probability of 0.1 to our training data in order to alleviate the overfitting issue. Specifically, the affine matrix used to augment the training data is defined as: $A = R_x R_y R_z T_{xyz}$, where R_x, R_y and R_z denotes the rotation matrices over x, y and z axis respectively. T_{xyz} represents the composition of a translation and scaling matrices. We randomly sample the degree of rotation, translation and scaling parameters from $[-10, 10]$, $[-0.05, 0.05]$ and $[-0, 08, 0.08]$.

3 Results

We implemented our method in Pytorch [13] and trained it on an NVIDIA RTX2080TI GPU using the Learn2Reg 2020 training dataset. As our method is not capable of multi-modal registration, we use a conventional affine registration approach (6 DoF) with the Normalized gradient field as the similarity metric and the Limited-memory the L-BFGS as the optimizer for task 1. Table 1 shows the results of our model and the other top-performance teams on the Learn2Reg 2020 testing dataset. Our method achieves the best overall ranking among ten international groups. In particular, our method shows the state-of-the-art results in abdominal CT registration (Task 3) and Hippocampus MR registration (Task 4) in terms of the dice similarity coefficient, Hausdorff distance of segmentation, smoothness of the solutions and runtime. Qualitative results of abdominal CT registration and Hippocampus MR registration on cases with large inter-variation are depicted in Figs. 2 and 3. The result in Fig. 2 shows that not only does our method achieve promising registration accuracy on the labeled anatomical structures under the large deformation setting, our method is also capable to register the unlabeled regions, including spinal vertebra and pelvis.

Table 1. Learn2Reg 2020 testing dataset results. Average target registration error of landmarks (TRE), Dice score (DSC), Standard deviation of log Jacobian determinant of the deformation field (SDlogJ) and runtime in second(s) (Time) for each task.

Method	Task1			Task2			Task3			Task4		
	TRE	SDlogJ	Time	TRE	SDlogJ	Time	DSC	SDlogJ	Time	DSC	SDlogJ	Time
Initial	6.38	-	-	10.24	-	-	0.23	-	-	0.55	-	-
Ours	5.67	0.00	31.21	3.24	0.06	1.33	0.67	0.12	1.83	0.88	0.05	1.00
PDD-Net [5,6]	3.08	0.00	4.61	2.46	0.04	2.49	0.46	0.43	4.05	0.78	0.07	0.31
deeds [7,8]	3.93	0.00	9.12	2.26	0.07	41.32	0.51	0.11	41.65	0.76	0.11	3.14

4 Conclusion

In this paper, we have described the method that we used in the Learn2Reg 2020 challenge. By integrating the conventional multi-resolution optimization strategy with deep neural network, our method inherits the runtime advantage of the deep neural network and achieves the state-of-the-art results in most of the clinically relevant sub-tasks in the Learn2Reg 2020 challenge.

References

1. Learn2reg: 2020 miccai registration challenge. https://learn2reg.grand-challenge.org/. Accessed 10 Nov 2020

2. Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V.: An unsupervised learning model for deformable medical image registration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9252–9260 (2018)
3. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018, Part I. LNCS, vol. 11070, pp. 729–738. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_82
4. Ghiasi, G., Fowlkes, C.C.: Laplacian pyramid reconstruction and refinement for semantic segmentation. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016, Part III. LNCS, vol. 9907, pp. 519–534. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46487-9_32
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, Part III. 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, M., Schnabel, J.A.: MRF-based deformable registration and ventilation estimation of lung CT. IEEE Trans. Med. Imaging **32**(7), 1239–1248 (2013)
8. Heinrich, M.P., Maier, O., Handels, H.: Multi-modal multi-atlas segmentation using discrete optimisation and self-similarities. VISCERAL Challenge@ ISBI 1390, 27 (2015)
9. Hering, A., Murphy, K., Ginneken, B.V.: Lean2Regchallenge: CT lung registration-training data [data set]. Zenodo (2020)
10. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Fast and accurate image super-resolution with deep Laplacian pyramid networks. IEEE Trans. Pattern Anal. Mach. Intell. **41**(11), 2599–2613 (2018)
11. Mok, T.C., Chung, A.: Fast symmetric diffeomorphic image registration with convolutional neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4644–4653 (2020)
12. Mok, T.C.W., Chung, A.C.S.: Large deformation diffeomorphic image registration with Laplacian pyramid networks. In: Martel, A.L., et al. (eds.) MICCAI 2020, Part III. LNCS, vol. 12263, pp. 211–221. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59716-0_21
13. Paszke, A., Gross, S., Chintala, S., et al.: Automatic differentiation in pytorch. In: NIPS-W (2017)
14. Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., Pennec, X.: SVF-Net: learning deformable image registration using shape matching. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D.L., Duchesne, S. (eds.) MICCAI 2017. LNCS, vol. 10433, pp. 266–274. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66182-7_31
15. Simpson, A.L., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019)
16. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Staring, M., Išgum, I.: End-to-end unsupervised deformable image registration with a convolutional neural network. In: Cardoso, M., et al. (eds.) DLMIA/ML-CDS -2017. LNCS, vol. 10553, pp. 204–212. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67558-9_24

17. Xiao, Y., Fortin, M., Unsgård, G., Rivaz, H., Reinertsen, I.: REtrospective evaluation of cerebral tumors (resect): a clinical database of pre-operative MRI and intra-operative ultrasound in low-grade glioma surgeries. Med. Phys. **44**(7), 3875–3882 (2017)
18. Xu, Z., et al.: Evaluation of six registration methods for the human abdomen on clinically acquired CT. IEEE Trans. Biomed. Eng. **63**(8), 1563–1572 (2016)
19. Zhao, S., Dong, Y., Chang, E.I., Xu, Y., et al.: Recursive cascaded networks for unsupervised medical image registration. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 10600–10610 (2019)