Deep Learning-based Longitudinal Intra-subject Registration of Pediatric Brain MR Images *

Andjela Dimitrijevic[1,2], Vincent Noblet[3], and Benjamin De Leener[1,2,4]

[1] NeuroPoly Lab, Institute of Biomedical Engineering, Polytechnique Montréal, Montréal, QC, Canada
[2] Research Center, Ste-Justine Hospital University Centre, Montréal, QC, Canada
[3] ICube-UMR 7357, Université de Strasbourg, CNRS, Strasbourg, France
[4] Computer Engineering and Software Engineering, Polytechnique Montréal, Montréal, QC, Canada

Abstract. Deep learning (DL) techniques have the potential of allowing fast deformable registration tasks. Studies around registration often focus on adult populations, while there is a need for pediatric research where less data and studies are being produced. In this work, we investigate the potential of unsupervised DL-based registration in the context of longitudinal intra-subject registration on 434 pairs of publicly available Calgary Preschool dataset of children aged 2-7 years. This deformable registration task was implemented using the DeepReg toolkit. It was tested in terms of input spatial image resolution (1.5 vs 2.0 mm isotropic) and three pre-alignement strategies: without (NR), with rigid (RR) and with rigid-affine (RAR) initializations. The evaluation compares regions of overlap between warped and original tissue segmentations using the Dice score. As expected, RAR with an input spatial resolution of 1.5 mm shows the best performances. Indeed, RAR has an average Dice score of of 0.937±0.034 for white matter (WM) and 0.959±0.020 for gray matter (GM) as well as showing small median percentages of negative Jacobian determinant (JD) values. Hence, this shows promising performances in the pediatric context including potential neurodevelopmental studies.

Keywords: Learning-based image registration · pediatric · MRI.

1 Introduction

Registration consists of bringing a pair of images into spatial correspondence. There are hardly any registration methods dedicated to the pediatric brain, mainly because of the difficulties arising from major changes that occur during neurodevelopment [4]. Conventional deformable registration involves estimating a deformation field through an iterative optimization problem. This process is time consuming, but provides accurate results. Convolutional neural networks (CNN) can allow faster registrations by applying a learning-based approach [3].

* Supported by Polytechnique Montréal, by the Canada First Research Excellence Fund, and by the TransMedTech Institute.

Hence, applying DL methods to pediatric brain scans could improve registration and future diagnostics for medical applications. Ultimately, it would be relevant to validate the potential use of DL-based frameworks for pediatric populations.

The general objective of this study is to validate a DL framework which allows fast intra-subject deformable registrations after training on pediatric MRI scans. To do so, different initial conditions are considered by fragmenting the non-rigid transformation into its simpler parts. Pre-network rigid registration, RigidReg (RR) and rigid-affine registration, RigidAffineReg (RAR) are performed on each intra-subject pair using ANTs [1] in order to determine their respective impact on the network's performance. Also, a third method called NoReg (NR) is investigated where no pre-alignment task is done. These three methods are then trained using a U-Net like CNN architecture implemented via the DeepReg toolkit [6]. The robustness of these DL techniques is assessed by using different input resolutions (1.5 vs 2.0 mm isotropic) for the same network architectures.

2 Methodology

Preprocessing Pipeline. Each image was corrected for bias field inhomogeneity using N4 algorithm. Both rigid and rigid-affine pre-alignments were performed with ANTs registration framework. The Mattes similarity metric was used.

Unsupervised Deformable Registration Framework. The U-Net architecture used to generate the deformation field consists of a 3-layer encoder and decoder with 8, 16 and 32 channels each. As for the loss function, it is composed of a local normalized cross-correlation similarity measure and an L2-norm gradient regularization factor to ensure realistic physical deformation fields. Local normalized cross-correlation is chosen for its robustness to local variations of intensities. The ADAM optimizer is used with a learning rate set to 1.0e-4. Finally, the network was trained on a GeForce RTX 2080 Ti GPU.

3 Experiments

Data. 434 pairs of moving/fixed 3D images were extracted from the longitudinal Calgary Preschool dataset [5] containing 247 T1-weighted images from 64 children aged 2-7 years old. The average time interval between consecutive scans is of 1.15±0.68 years. The original images have a native resolution of 0.4492x0.4492x0.9 mm^3. The resized images of 1.5 mm as well as 2.0 mm isotropic resolution have respectively a matrix size of 153x153x125 and 114x114x94.

Evaluation. To acquire white matter (WM), gray matter (GM) and cerebrospinal fluid (CSF) segmentations for evaluation purposes, each image was non-linearly registered to the MNI pediatric template for children 4.5–8.5 years old [2]. This also allowed obtaining skull-stripped images via the available mask in the template space. Unsupervised networks are then evaluated using the Dice score as a performance metric. In addition, the generated deformation fields are

evaluated using the percentage of negative JD values indicating unwanted local foldings. To compare the impact from the three initialization methods or input resolutions, one-sided Wilcoxon signed-rank tests were performed.

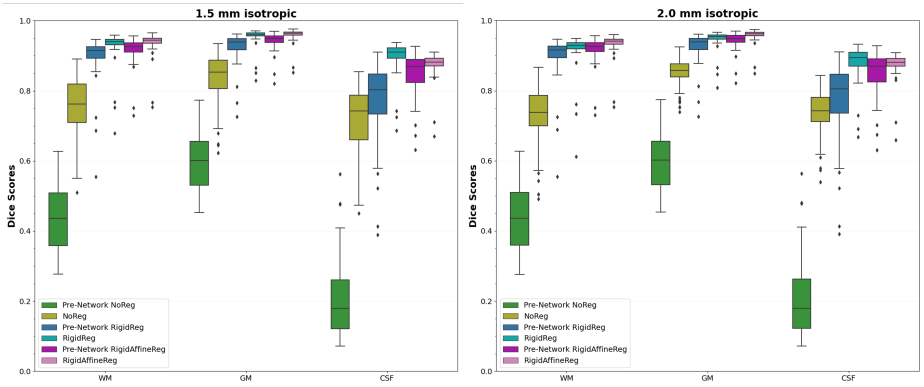

Fig. 1. Dice scores results for different input resolutions obtained for each method compared to their pre-network Dice scores represented as boxplots. The Dice scores are calculated for all subjects and WM, GM and CSF regions using the test set.

Table 1. Average Dice scores per resolution calculated over all segmented regions and subjects using the test set for the three studied methods. Median percentages of negative JD values are given because of highly right-skewed distributed data. ANTs pre-registration tasks are performed on the native resolution of 0.4492x0.4492x0.9 mm^3 and using the available CPU implementation.

	1.5 mm isotropic				2.0 mm isotropic				Native resolution
Methods	**Dice score**	**% of JD<0**	**Train time/epoch**	**Test time/pair**	**Dice score**	**% of JD<0**	**Train time/epoch**	**Test time/pair**	**ANTs pre-reg time/pair**
NR	0.764 ± 0.105	1.11e-1	189.3s	4.88s	0.770 ± 0.088	1.33e-1	74.8s	1.87s	0s
RR	0.929 ± 0.045	1.86e-4	137.7s	3.55s	0.916 ± 0.051	0	78.1s	1.88s	168.6s
RAR	0.924 ± 0.047	0	177.5s	4.13s	0.922 ± 0.047	0	75.8s	1.86s	365.8s

4 Results

A 85/15 % split was respectively done for train and test sets. Then, a three-fold cross-validation technique is employed to train and evaluate each method containing 123 pairs per fold. In total, 65 pairs are used for test purposes. Above, presented results for the two evaluated resolutions in Table 1 and Fig. 1 come from this unseen test set. Fig. 2 shows the differences of obtained predicted fixed images for all the three considered initialization methods. Also, two one-sided Wilcoxon tests were conducted comparing, first, the median Dice scores differences between RR and RAR (RAR-RR) for both resolutions as well as between 2.0 and 1.5 (1.5-2.0) for all initialization methods. The first test allowed rejecting the null hypothesis only for WM and GM ($p<1.06e-10$) showing higher median Dice scores for RAR compared to RR for both resolutions. The second

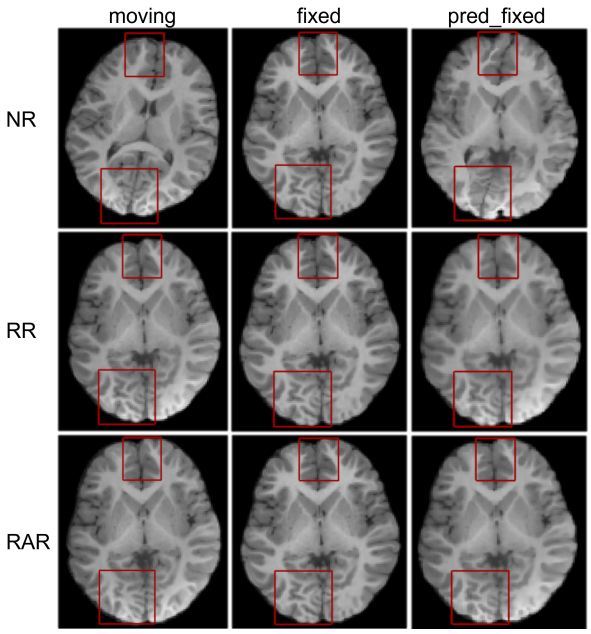

Fig. 2. Resulting images for a specific pair (age interval of 3.37 years) for the three pre-alignment strategies using an input spatial image resolution of 1.5 mm isotropic.

test shows that 1.5 mm isotropic resolution, at the cost of longer train and test times, yields slightly, but statistically significant better performances than 2.0 for RR and RAR methods for all segmented regions ($p<1.51e-4$). This improvement is not significant for GM and CSF regions for the NR method.

5 Discussion & Conclusion

In this study, we demonstrated that DL-based deformable registration succeeds to improve registration accuracy regardless of the initialization method and for both tested resolutions (see Fig. 1). RAR demonstrated higher Dice scores compared to RR for WM (0.937 ± 0.034 vs 0.930 ± 0.046) and GM (0.959 ± 0.020 vs 0.955 ± 0.025). Differing results for CSF may be due to its thin surface and errors arising from the skull-stripping process. Both RR and RAR reached high registration quality, while NR shows lower registration performance due to its incapacity to extract both global and local transformations simultaneously, shown in Fig. 2. However, NR remains relevant as no prior registration is needed. Future work will evaluate the capacity of a neural network to decompose the global and local transformations. Considering all studied combinations of pre-alignment strategies and input resolutions, RAR provides better Dice scores with 1.5 mm isotropic resolution images, which could help to perceive neurodevelopmental changes from a large age range of pediatric data.

References

1. Avants, B.B., et al.: A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage **54**(3), 2033–2044 (Feb 2011). https://doi.org/10.1016/j.media.2007.06.004
2. Fonov, V.S., et al.: Unbiased average age-appropriate atlases for pediatric studies. Neuroimage **54**(1), 313–27 (2011). https://doi.org/10.1016/j.neuroimage.2010.07.033
3. Haskins, G., et al.: Deep learning in medical image registration: a survey. Machine Vision and Applications **31**(1-2) (2020). https://doi.org/10.1007/s00138-020-01060-x
4. Phan, T.V., et al.: Processing of structural neuroimaging data in young children: Bridging the gap between current practice and state-of-the-art methods. Dev Cogn Neurosci **33**, 206–223 (2018). https://doi.org/10.1016/j.dcn.2017.08.009
5. Reynolds, J.E., et al.: Calgary preschool magnetic resonance imaging (mri) dataset. Data Brief **29**, 105224 (2020). https://doi.org/10.1016/j.dib.2020.105224
6. Yunguan, F., et al.: Deepreg: a deep learning toolkit for medical image registration. Journal of Open Source Software **5**(55), 2705 (2020). https://doi.org/10.21105/joss.02705