Attention-Driven Multi-channel Deformable Registration of Structural and Microstructural Neonatal Data

Irina Grigorescu[1,2(✉)], Alena Uus[1,2], Daan Christiaens[1,3], Lucilio Cordero-Grande[1,2,5], Jana Hutter[1], Dafnis Batalle[1,4], A. David Edwards[1], Joseph V. Hajnal[1,2], Marc Modat[2], and Maria Deprez[1,2]

[1] Centre for the Developing Brain, School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK
irina.grigorescu@kcl.ac.uk
[2] Biomedical Engineering Department, School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK
[3] Departments of Electrical Engineering, ESAT/PSI, KU Leuven, Leuven, Belgium
[4] Department of Forensic and Neurodevelopmental Science, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[5] Biomedical Image Technologies, ETSI Telecomunicación, Universidad Politécnica de Madrid & CIBER-BNN, Madrid, Spain

Abstract. Image registration of structural and microstructural data allows accurate alignment of anatomical and diffusion channels. However, existing techniques employ simple fusion-based approaches, which use a global weight for each modality, or empirically-driven approaches, which rely on pre-calculated local certainty maps. Here, we present a novel attention-based deep learning deformable image registration solution for aligning multi-channel neonatal MRI data. We learn optimal attention maps to weigh each modality-specific velocity field in a spatially varying fashion, thus allowing for local fusion of structural and microstructural images. We evaluate our proposed method on registrations of 30 multi-channel neonatal MRI to a standard structural and microstructural atlas, and compare it against models trained without the use of attention maps on either single or both modalities. We show that by combining the two channels through attention-driven image registration, we take full advantage of the two complementary modalities, and achieve the best overall alignment of both structural and microstructural data.

Keywords: multi-channel registration · attention maps · deep learning registration

1 Introduction

The neonatal brain undergoes dramatic changes during early life, such as cortical folding and myelination. Non-invasive magnetic resonance imaging (MRI) offers

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
R. Licandro et al. (Eds.): PIPPI 2022, LNCS 13575, pp. 71–81, 2022.
https://doi.org/10.1007/978-3-031-17117-8_7

snapshots of the evolving morphology and tissue properties in developing brain across multiple subjects and time-points. As a prerequisite of further analysis, MRI of various modalities needs to be aligned. Structural and microstructural MRI modalities offer complementary information about morphology and tissue properties of the developing brain, however inter-subject alignment is most commonly driven by a single modality (structural [2] or diffusion [23]). Studies have shown that combining diffusion and structural data to drive the registration [1,7,8,20] improves the overall alignment. Classic approaches for fusing these channels are based on simple averaging of the deformation fields from the individual channels [1], or weighting the deformation fields based on certainty maps calculated from normalised gradients correlated to structural content [7,19,20].

In order to establish accurate correspondences between MR images acquired during the neonatal period, we propose an attention-driven multi-channel deep learning image registration framework that aims to combine information from T_2-weighted (T_2w) neonatal scans with diffusion weighted imaging (DWI)-derived fractional anisotropy (FA) maps. Our proposed solution selects the most salient features from these 2 image modalities to improve alignment of individual MRI images to a common atlas space.

More specifically, we train conditional variational autoencoder (CVAE) image registration networks to align either structural or microstructural data to 36 weeks neonatal atlas [19] of the same modality. As a second step, we build a convolutional neural network (CNN) which learns attention maps for weighted combination of the predicted modality-specific velocity fields to achieve an optimal multi-channel alignment. Throughout this work, we use 3-D MRI brain scans [6] acquired as part of the developing Human Connectome Project (dHCP[1] as the moving images, and 36 weeks neonatal multi-modal atlas[2] [19] as the fixed image.

We evaluate our proposed framework on a test set of 30 neonates scanned around 40 weeks post-menstrual age (PMA), and we compare the results against registration networks trained on T_2w-only, FA-only, and both modalities at the same time, either without attention, or with previously proposed attention mechanism [9,21]. The quantitative evaluation confirmed that while cortical structures were better aligned using T_2w data and white matter tracts were better aligned using FA maps, the attention-based multi-channel registration aligned both types of structures accurately.

2 Method

Image Registration Network. In this study, we employ a CVAE [11] to model the registration probabilistically as proposed by [12]. In short, a pair of 3D MRI volumes M_{T2w} and F_{T2w} (or M_{FA} and F_{FA}) are passed through the network to learn a velocity field v_{T2w} (or v_{FA}). The *exponentiation layers* (with 4 *scaling-and-squaring* [3] steps) transform it into a topology-preserving deformation field

[1] developingconnectome.org.

[2] gin.g-node.org/alenaullauus/4d_multi-channel_neonatal_brain_mri_atlas.

ϕ_{T2w} (or ϕ_{FA}). A *Spatial Transformer* layer [5] is then used to warp (linearly resample) the moving images M_{T2w} (or M_{FA}) and obtain the moved image $M_{T2w}(\phi_{T2w})$ (or $M_{FA}(\phi_{FA})$). We keep the network architecture similar to the original paper [12], but use a latent code size of 32 and a Gaussian smoothing layer with $\sigma = 1$ mm (kernel size 3^3). Throughout this work, we use 36 weeks old neonatal structural (T_2w) and microstructural (FA maps) atlases [19] as the fixed images. We have chosen this age for the templates due to the lower degree of gyrification which facilitates a more accurate registration of the cortex across the cohort.

Attention Image Registration Network. We construct a CNN which uses pairs of modality-specific velocity fields as an input, and outputs a combined velocity field which aims to align both structural and microstructural data simultaneously. The network learns the attention maps α_{T2w} and α_{FA}, for which $\alpha_{T2w} + \alpha_{FA} = 1$ at every voxel. The input velocity fields are weighted with the attention maps and combined to create a final velocity field v.

The architecture of our proposed *attention image registration network* is presented in Fig. 1. For each subject in our dataset, we employ the previously trained *registration-only networks* on either pairs of T_2w images (M_{T2w} and F_{T2w}) or FA maps (M_{FA} and F_{FA}) to output modality-specific velocity fields v_{T2w} and v_{FA}. These two fields are concatenated and put through three 3D convolutional layers (stride 2) of 16, 32, and 64 filters, respectively, with a kernel size of 3^3, followed by *Leaky ReLU* ($\alpha = 0.2$) activations [22]. The activation maps of the final layer are concatenated with the subject's moving images M_{T2w} and M_{FA} downsampled to size 16^3. This is followed by three 3D convolutional layers (stride 1) of 32, 16, and 16 filters, respectively, with a kernel size of 3^3, *Leaky ReLU* ($\alpha = 0.2$) activations and upsampling. The final two layers are: one 3D convolutional layer (with stride 1, 8 filters, and *Leaky ReLU* activation), and one 3D convolutional layer (with stride 1, and 2 filters), followed by a *Softmax* activation function which outputs the two modality-specific attention maps α_{T2w} and α_{FA}. The final velocity field is created as $v = v_{T2w} \odot \alpha_{T2w} + v_{FA} \odot \alpha_{FA}$, where $\odot$ represents element-wise multiplication. Similar to the registration network, the velocity field v is put through an *exponentiation layer* to create the combined field ϕ, which is then used to warp the moving volumes M_{T2w} and M_{FA}.

Channel and Spatial Attention Network. To compare our proposed attention-driven image registration network with other attention techniques, we add channel and spatial attention modules throughout the image registration network. More specifically, after every convolutional layer of the network, we add a channel attention module (squeeze-and-excitation block [9]), followed by a spatial attention module [21]. In total, we add 4 channel+spatial attention modules in the encoder part of the CVAE, and 5 modules in the decoder.

Loss Functions. For this study, we train the *registration-only network* and the *channel+spatial attention network* using the following loss function:

$$\mathcal{L}_{reg} = \mathcal{L}_{KLD} + \lambda \left(\lambda_{T2w} \mathcal{L}_{NCC}^{T2w} + \lambda_{FA} \mathcal{L}_{NCC}^{FA} \right) + \lambda_{reg} \mathcal{L}_{BE} \tag{1}$$

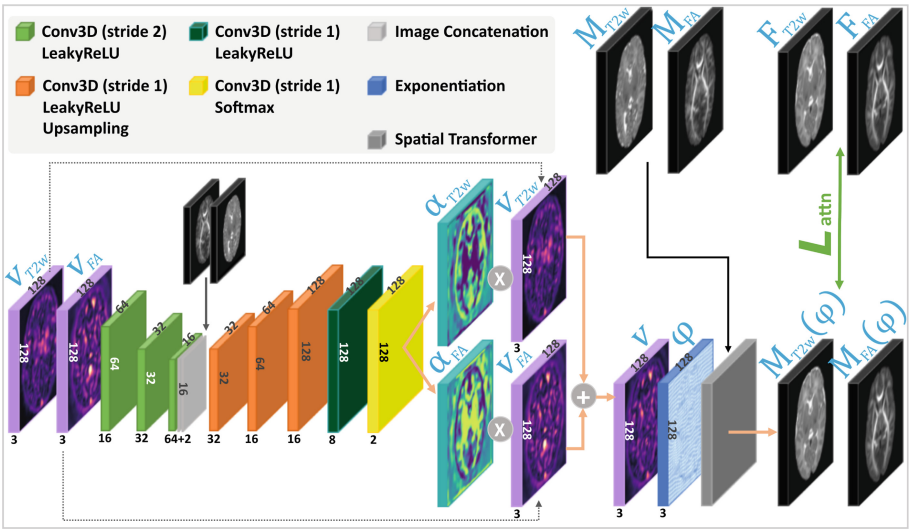

Fig. 1. Our proposed attention-based image registration network architecture, which uses as input subject- and modality-specific velocity fields (v_{T2w} and v_{FA}). The attention network outputs two 1-channel maps α_{T2w} and α_{FA} which are used to create a combined velocity field v. The velocity field v is transformed into a dense displacement field ϕ which warps the subject's moving images (M_{T2w} and M_{FA}) into $M_{T2w}(\phi)$ and $M_{FA}(\phi)$. The network is trained to achieve good alignment between the warped images and the fixed atlases (F_{T2w} and F_{FA}).

and our proposed *attention network* with:

$$\mathcal{L}_{attn} = \lambda_{T2w}\,\mathcal{L}_{NCC}^{T2w} + \lambda_{FA}\,\mathcal{L}_{NCC}^{FA} \tag{2}$$

where λ, λ_{reg}, λ_{T2w} and λ_{FA} are hyperparameters, $\mathcal{L}_{KLD}$ is the Kullback-Leibler (KL) divergence, $\mathcal{L}_{NCC}$ is the global symmetric normalised cross correlation (NCC) dissimilarity measure, and $\mathcal{L}_{BE}$ is a bending energy regularisation penalty [16]. In this study, we set $\lambda_{reg} = 0.01$ and $\lambda = 5000$ (as proposed in [12]).

Training. First, using the no-attention *registration-only network*, we train 2 single-modality models on either pairs of T_2w-only data ($\lambda_{T2w} = 1.0$, $\lambda_{FA} = 0.0$) or FA-only data ($\lambda_{T2w} = 0.0$, $\lambda_{FA} = 1.0$). Then, we train the three networks (the no-attention *registration-only network*, the *channel+spatial attention network*, and our proposed *attention network*) on both modalities, using the following sets of hyperparameters: $(\lambda_{T2w}, \lambda_{FA}) = \{(1.0, 0.1), (1.0, 0.175), (1.0, 0.25), (1.0, 0.5), (1.0, 0.75), (1.0, 1.0)\}$. In total, we have 20 models: 8 using the *registration-only network*, 6 using the *channel+spatial attention network*, and 6 with our proposed *attention network*.

We train the 20 models until convergence (150 epochs, or 52500 iterations), using the Adam optimizer with its default parameters ($\beta_1 = .9$ and $\beta_2 = .999$), a decaying cyclical learning rate scheduler [17] with a base learning rate of 10^{-6}

and a maximum learning rate of 10^{-3}, and an L_2 weight decay factor of 10^{-5}. All networks were implemented in PyTorch (v1.10.2), with TorchIO (v0.18.73) [15] for data preprocessing (intensity normalisation) and loading, and training was performed on a 12 GB Titan XP. Average inference times were: 0.16 s/sample for the *registration-only networks*, 0.31 s/sample for the *attention-based networks*, and 0.63 s/sample for the *channel+spatial attention networks*.

3 Results

Image Selection and Preprocessing For this study, we use a total of 414 T_2w images and FA maps of neonates born between 23–42 weeks gestational age (GA) and scanned at term-equivalent age (37–45 weeks PMA) [6]. As preprocessing steps, we first affinely pre-registered the data to a common 36 weeks gestational age atlas space [19] using the MIRTK software toolbox [16], and then we resampled both structural and microstructural volumes to be 1 mm isotropic resolution. To obtain the FA maps, we used the MRtrix3 toolbox [18], and we performed skull-stripping using the available dHCP brain masks [4]. Finally, we cropped the resulting images to a $128 \times 128 \times 128$ size.

Out of the 414 subjects in our dataset, we used 350 for training, 34 for validation and 30 subjects for test, as described in Table 1. We used the validation set to inform us about our models' performance during training, and we report all of our results on the test set.

Table 1. Number of scans in different datasets used for training, validation and testing the models, together with their mean GA at birth (standard deviation) and mean PMA at scan (standard deviation).

Dataset	**#Subjects**	**GA** [weeks]	**PMA** [weeks]
Train	350 (164♀ + 186♂)	38.0 (3.8)	40.6 (1.9)
Validate	34 (14♀ + 20♂)	39.7 (1.4)	40.7 (1.7)
Test	30 (12♀ + 18♂)	39.8 (1.5)	40.6 (1.9)

Quantitative Evaluation. To validate which of the 20 models performs best, we carry out a quantitative evaluation on our test dataset of 30 subjects. Each subject and the atlas had the following tissue label segmentations obtained from T_2w images using the Draw-EM pipeline [14]: cortical gray matter (cGM), white matter (WM), ventricles, hippocampi and amygdala. Additionally, a WM structure called the internal capsule (IC) was manually segmented on FA maps of all test subjects. These labels were propagated from each subject into the atlas space using the predicted deformation fields. To evaluate performance of the registration, Dice scores and average surface distances (SimpleITK v2.1.1 [13]) were calculated between the warped labels and the atlas labels.

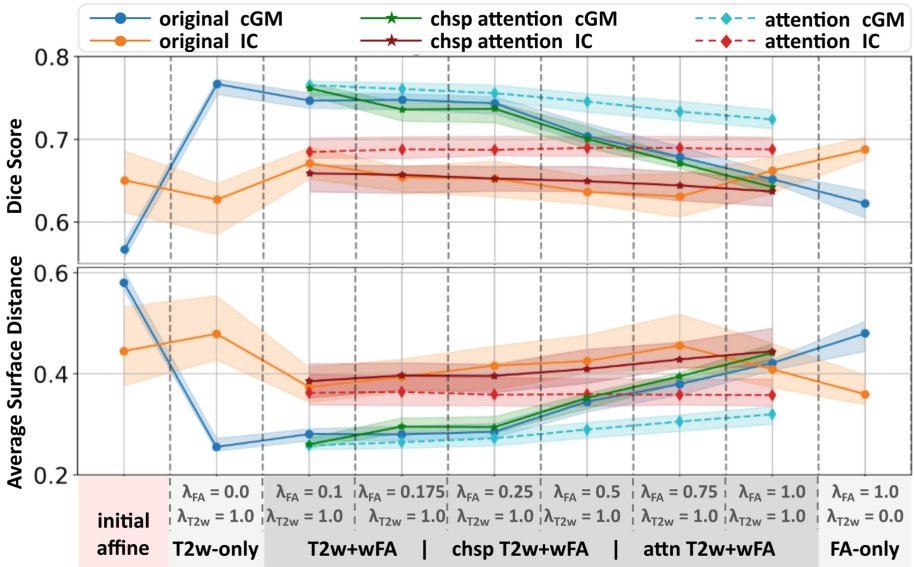

Fig. 2. Line plots showing median Dice scores (first row) and average surface distances (second row) for cGM and IC structures, with the first column showing their initial affine alignment. The dark blue lines (*original cGM*) and the orange lines (*original IC*) show the scores for the registration models without attention. The light blue (*attention cGM*) and the red (*attention IC*) plots represent the values obtained by our *proposed attention-driven image registration network*, while the green (*chsp attention cGM*) and the brown (*chsp attention IC*) lines represent the values obtained by the *channel+spatial attention network*, for different values of the λ_{T2w} and λ_{FA} hyperparameters. The shading around each median line is the IQR. (Color figure online)

First, we looked at how the models performed based on two tissue types (the cGM and the IC). We chose these structures because the cGM delineation is poor on the FA maps, while the IC is a white matter structure which is very prominent in the microstructure data. Both Dice scores and average surface distances are summarised in Fig. 2, where the first column shows the values for the initial affine alignment, while the second and last columns show the T_2w-only and the FA-only image registration networks. Columns 3–8 show different multi-channel models for increasing values of the λ_{FA} hyperparameter, while λ_{T2w} is kept the same.

The best overall performance in terms of Dice scores and average surface distances is obtained by our *proposed attention model* for $\lambda_{T2w} = 1.0$ and $\lambda_{FA} = 0.1$ (third column, Fig. 2), where the cGM is aligned as well as the T_2w-*only* model, and the IC structure as good as the FA-only model (the differences are not statistically significant). Using *channel+spatial attention* with the same hyperparameter setup ($\lambda_{T2w} = 1.0$ and $\lambda_{FA} = 0.1$) achieves good results for the cGM structure, but cannot align the IC structure as well as the FA-only model, or the *proposed attention model*.

For the T_2w-only model (second column) the IC is poorly aligned, obtaining scores which are worse than the initial affine alignment, while the cGM label obtains the best alignment. On the other hand, for the FA-only model (last column) the IC is well aligned, while the cGM obtains lower scores. In the original registration networks (dark blue and orange) we see a steady worsening of cGM scores as λ_{FA} increases, while the IC structure varies across the different λ_{FA} values. For the attention-driven networks (light blue and red), the scores in cGM degrade more gently, while the IC structures remain steady. Finally, the proposed attention networks always outperform the multi-channel registration networks with no attention, and this improvement is statistically significant for all values of λ_{FA}.

Table 2 shows the results of 6 of our trained models for all tissue types (cGM, WM, ventricles, hippocampi and amygdala, and IC). Here, we call the $T_2\text{w} + \boldsymbol{w}FA$, the *chsp* $T_2\text{w} + \boldsymbol{w}FA$, and the *attn* $T_2\text{w} + \boldsymbol{w}FA$ models as the ones trained with the lowest weight on the FA maps ($\lambda_{T2w} = 1.0$ and $\lambda_{FA} = 0.1$).

Table 2. Mean (±standard deviation) Dice scores (DS) and average surface distances (ASD) on test set. Best scores are highlighted in bold (*t-test* $p < 0.05$), while the green shading highlights the model which performed best amongst the ones which use both T_2w and FA modalities (*t-test* $p < 0.05$). The multi-modality weighted models shown here use $\lambda_{T2w} = 1.0$ and $\lambda_{FA} = 0.1$.

Model	cGM	WM	Ventricles	Amygdala	IC	
affine	.567±.02	.7±.03	.631±.05	.746±.05	.642±.07	DS
T2w-only	**.763±.01**	**.844±.02**	**.797±.02**	.803±.02	.614±.04	DS
FA-only	.621±.02	.756±.02	.676±.04	.769±.03	**.686±.03**	DS
T2w+FA	.653±.01	.766±.01	.742±.03	.782±.02	.655±.03	DS
T2w+**w**FA	.747±.01	.826±.02	.775±.02	.808±.02	.669±.03	DS
chsp T2w+**w**FA	.761±.01	.841±.01	.791±.01	**.814±.02**	.656±.03	DS
attn T2w+wFA	**.763±.01**	.842±.01	.793±.02	**.816±.02**	**.683±.03**	DS
affine	.582±.04	.409±.04	.508±.1	.31±.08	.479±.1	ASD
T2w-only	**.259±.02**	**.193±.02**	**.242±.05**	.233±.04	.498±.09	ASD
FA-only	.477±.04	.319±.02	.433±.09	.276±.05	**.374±.05**	ASD
T2w+FA	.419±.02	.317±.02	.324±.06	.266±.04	.417±.06	ASD
T2w+**w**FA	.279±.01	.218±.02	.264±.04	.223±.04	.383±.05	ASD
chsp T2w+**w**FA	**.262±.01**	.198±.01	.248±.04	**.209±.03**	.39±.05	ASD
attn T2w+wFA	**.260±.02**	.197±.01	.248±.04	**.212±.03**	**.37±.05**	ASD

Our proposed *attn* $T_2\text{w} + \boldsymbol{w}FA$ model has the best overall performance. For structures which were delineated in T_2w images, the proposed attention model performed better (hippocampi and amygdala), equally well (cGM), or very close (WM, ventricles) to the T_2w-*only* model, showing that thanks to attention we are able to keep advantages of structural only registration. For IC, which was

derived from FA maps, the proposed attention model performed equally well to the *FA-only* model, showing that the attention also allows us to keep the advantages of the microstructural only registration model.

Using *channel+spatial attention* helped with the alignment of the structural labels (cGM, WM, ventricles, hippocampi and amygdala), but had significantly lower performance for IC (lower than the no-attention T_2w+**w**FA model).

The T_2w-*only* model performed slightly worse for the hippocampi and amygdala, while the scores for the IC structure were worse than the initial affine alignment. The *FA-only* model obtains poor scores in all structures except the IC. Finally, the multi-channel models trained without attention always performed worse than the attention-driven models. In fact, the T_2w+*FA* network, where $\lambda_{T2w} = \lambda_{FA} = 1.0$, obtained the lowest performance amongst the multi-channel models, showing that besides attention, the global weighting ($\lambda_{FA} = 0.1$) was an important factor towards the network's performance.

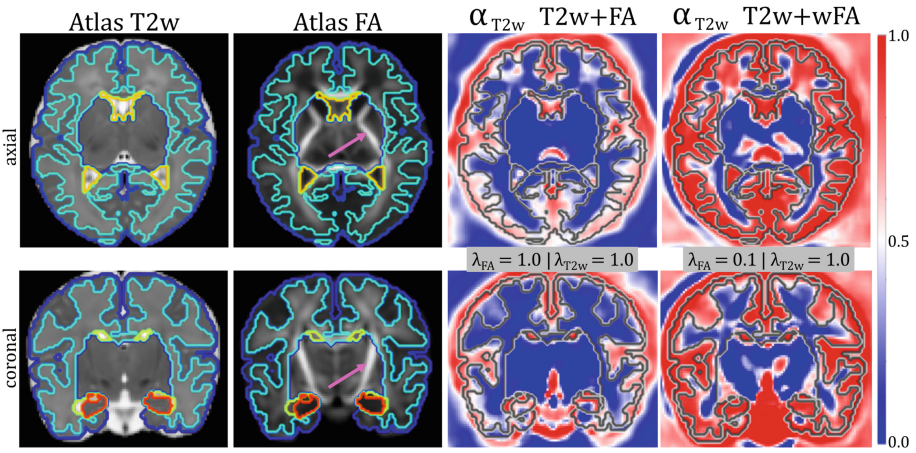

Fig. 3. Mid-brain axial and coronal slices of both T_2w and FA fixed images (first two columns), together with average α_{T2w} attention maps for the *attn* T_2w+*FA* with $\lambda_{FA} = \lambda_{T2w} = 1.0$, and the *attn* T_2w + $\boldsymbol{w}$*FA* with $\lambda_{T2w} = 1.0$, $\lambda_{FA} = 0.1$ models on the last two columns. Contour lines of the boundaries between cGM (dark blue), WM (cyan), ventricles (yellow) and hippocampi and amygdala (red) are overlaid on top, while the pink arrow points to the IC structure. (Color figure online)

Visualisation of Attention Maps. Figure 3 shows average attention maps from 10 neonatal subjects scanned around 40 weeks PMA for two of our attention-driven models: *attn* T_2w+*FA* ($\lambda_{FA} = \lambda_{T2w} = 1.0$) and *attn* T_2w+$\boldsymbol{w}$*FA* ($\lambda_{T2w} = 1.0$, $\lambda_{FA} = 0.1$). The first two columns show the middle axial and coronal slices of the T_2w and FA atlases which were used for training, together with segmentation of the investigated brain structures. The last two columns show the average α_{T2w} attention maps (in atlas space) for the 2 models. We can observe that the α_{T2w} attention maps cover the cGM region, and this is more

pronounced when λ_{FA} is decreased from 1.0 to 0.1. On the other hand, α_{T2w} is close to zero in the area of the main white matter tracts in both cases.

4 Conclusion

This paper presents a novel solution for multi-channel registration, which combines structural and microstructural MRI data based on learned spatially varying attention maps that optimise the multi-channel alignment. Our quantitative evaluation showed that the proposed attention-driven image registration network improves overall alignment when compared to models trained on multi-channel data, while maintaining the performance of the single-channel registration for the structures delineated on that channel. Moreover, using attention helps drive the registration to better alignment of tissue structures, but only our proposed model obtains results on par to using microstructural data only in terms of aligning white matter labels.

The main limitations of this work are: the use of a single latent code size and smoothing kernel, no comparison with classic multi-channel image registration tools [10,19], and a limited number of labels used for validation. Future work will focus on evaluating the effect of the latent code size and smoothing kernel on the predicted velocity fields, and exploring the use of older neonate atlases. Moreover, we aim to extend our attention-driven image registration network to incorporate higher-order data, such as diffusion tensor (DT) images [8].

Acknowledgements. This work was supported by the Academy of Medical Sciences Springboard Award [SBF004\1040], Medical Research Council (Grant no. [MR/K006355/1]), European Research Council under the European Union's Seventh Framework Programme [FP7/20072013]/ERC grant agreement no. 319456 dHCP project, the EPSRC Research Council as part of the EPSRC DTP (grant Ref: [EP/R513064/1]), the Wellcome/EPSRC Centre for Medical Engineering at King's College London [WT 203148/Z/16/Z], the NIHR Clinical Research Facility (CRF) at Guy's and St Thomas', and by the National Institute for Health Research Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London.

References

1. Avants, B., Duda, J.T., Zhang, H., Gee, J.C.: Multivariate normalization with symmetric diffeomorphisms for multivariate studies. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007. LNCS, vol. 4791, pp. 359–366. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75757-3_44
2. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. **12**(1), 26–41 (2008)
3. 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)

4. Christiaens, D., et al.: Scattered slice SHARD reconstruction for motion correction in multi-shell diffusion MRI. Neuroimage **225**, 117437 (2021)
5. 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. LNCS, vol. 11070, pp. 729–738. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_82
6. Edwards, A.D., et al.: The developing human connectome project neonatal data release. Front. Neurosci. **16**, 886772 (2022)
7. Forsberg, D., Rathi, Y., Bouix, S., Wassermann, D., Knutsson, H., Westin, C.-F.: Improving registration using multi-channel diffeomorphic demons combined with certainty maps. In: Liu, T., Shen, D., Ibanez, L., Tao, X. (eds.) MBIA 2011. LNCS, vol. 7012, pp. 19–26. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24446-9_3
8. Grigorescu, I., et al.: Diffusion tensor driven image registration: a deep learning approach. In: Špiclin, Ž, McClelland, J., Kybic, J., Goksel, O. (eds.) WBIR 2020. LNCS, vol. 12120, pp. 131–140. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-50120-4_13
9. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
10. Irfanoglu, M.O., et al.: DR-TAMAS: diffeomorphic registration for tensor accurate alignment of anatomical structures. Neuroimage **132**, 439–454 (2016)
11. Kingma, D.P., Rezende, D.J., Mohamed, S., Welling, M.: Semi-supervised learning with deep generative models (2014). arXiv:1406.5298
12. Krebs, J., Mansi, T., Mailhé, B., Ayache, N., Delingette, H.: Unsupervised probabilistic deformation modeling for robust diffeomorphic registration. In: Stoyanov, D., et al. (eds.) DLMIA/ML-CDS -2018. LNCS, vol. 11045, pp. 101–109. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00889-5_12
13. Lowekamp, B., Chen, D., Ibanez, L., Blezek, D.: The design of SimpleiTK. Front. Neuroinform. **7**, 45 (2013)
14. Makropoulos, A., et al.: Automatic whole brain MRI segmentation of the developing neonatal brain. IEEE Trans. Med. Imaging **33**(9), 1818–1831 (2014)
15. Pérez-García, F., Sparks, R., Ourselin, S.: TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput. Methods Programs Biomed. **208**, 106236 (2021)
16. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans. Med. Imaging **18**(8), 712–721 (1999)
17. Smith, L.N.: Cyclical learning rates for training neural networks (2015)
18. Tournier, J.D., et al.: MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. Neuroimage **202**, 116137 (2019)
19. Uus, A., et al.: Multi-channel 4D parametrized atlas of macro- and microstructural neonatal brain development. Front. Neurosci. **15**, 721 (2021)
20. Uus, A., et al.: Multi-channel registration for diffusion MRI: longitudinal analysis for the neonatal brain. In: Špiclin, Ž, McClelland, J., Kybic, J., Goksel, O. (eds.) WBIR 2020. LNCS, vol. 12120, pp. 111–121. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-50120-4_11
21. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018.

LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
22. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network (2015). arXiv:1505.00853
23. Zhang, H., Yushkevich, P.A., Alexander, D.C., Gee, J.C.: Deformable registration of diffusion tensor MR images with explicit orientation optimization. Med. Image Anal. **10**(5), 764–785 (2006). The Eighth International Conference on Medical Imaging and Computer Assisted Intervention - MICCAI 2005