Uncertainty-Aware Deep Learning Based Deformable Registration

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 and CIBER-BNN, Madrid, Spain

Abstract. We introduce an uncertainty-aware deep learning deformable image registration solution for magnetic resonance imaging multi-channel data. In our proposed framework, the contributions of structural and microstructural data to the displacement field are weighted with spatially varying certainty maps. We produce certainty maps by employing a conditional variational autoencoder image registration network, which enables us to generate uncertainty maps in the deformation field itself. Our approach is quantitatively evaluated on pairwise registrations of 36 neonates to a standard structural and/or microstructural template, and compared with models trained on either single modality, or both modalities together. Our results show that by incorporating uncertainty while fusing the two modalities, we achieve superior alignment in cortical gray matter and white matter regions, while also achieving a good alignment of the white matter tracts. In addition, for each of our trained models, we show examples of average uncertainty maps calculated for 10 neonates scanned at 40 weeks post-menstrual age.

Keywords: Multi-channel registration · Uncertainty · Certainty maps

1 Introduction

Tracking changes in the developing brain depends on precise inter-subject image registration. However, most applications in this field rely on a single modality [3, 15], such as structural or diffusion data, to learn spatial correspondences between images, without taking into account the complementary information provided by using both. In general, T_2-weighted (T_2w) magnetic resonance imaging (MRI)

C. H. Sudre et al. (Eds.): UNSURE 2021/PIPPI 2021, LNCS 12959, pp. 54–63, 2021.
https://doi.org/10.1007/978-3-030-87735-4_6

scans have high contrast between different brain tissues and can delineate the cortical gray matter (cGM) region well, while diffusion weighted imaging (DWI) data primarily provides information on white matter (WM) structures.

Multi-channel registration which includes both structural and diffusion data has been shown to improve alignment [1,7] of images. However, one of the main challenges of this approach is the low contrast or homogeneous intensity which characterises different anatomical regions on both structural (*e.g.,* deep gray matter) and microstructural MRI (*e.g.,* cortex). Classic approaches for fusing these channels are based on simple averaging [1], or on calculating certainty maps based on normalised gradients correlated to structural content [7].

In order to establish accurate correspondences between MR images acquired during the neonatal period, we propose an uncertainty-aware deep learning image registration framework that allows local certainty-based fusion of T_2w neonatal scans with DWI-derived fractional anisotropy (FA) maps. More specifically, we employ a conditional variational autoencoder (CVAE) image registration network [11] and use it to calculate uncertainty maps in the generated dense displacement fields. We predict the displacement fields and their uncertainty for each individual modality, empirically calculate certainty maps and use them in a combined model which we call T_2w+FA+uncert.

Throughout this work we use 2-D MRI mid-brain axial slices acquired as part of the developing Human Connectome Project[1] (dHCP), and T_2w and FA 36 weeks gestational age templates [17] for the fixed slices. We showcase the capabilities of our proposed framework on images of infants born and scanned at different gestational ages, and we compare the results against three different models, trained on T_2w-only, on FA-only and on T_2w+FA scans. Our results show that in terms of Dice scores and average Hausdorff distances, our proposed model performs better in WM and cGM regions when compared to the T_2w+FA model, and better than the T_2w-only model in terms of aligning white matter structures (*i.e.,* internal capsule).

2 Method

Data Acquisition. The structural (T_2w) imaging data used in this study was acquired with a Philips Achieva 3T scanner and a 32-channels neonatal head coil [8], using a turbo spin echo (TSE) sequence ($T_R = 12$ s, echo time $T_E = 156$ ms, and SENSE factors of 2.11 for the axial plane and 2.58 for the sagittal plane). Images were acquired with an in-plane resolution of 0.8×0.8 mm, slice thickness of 1.6 mm and overlap of 0.8 mm. All data was motion corrected [5] and super-resolution reconstructed to a 0.5 mm isotropic resolution [12]. The DWI scans were acquired using a monopolar spin echo echo-planar imaging (SE-EPI) Stejksal-Tanner sequence [9]. A multiband factor of 4 and a total of 64 interleaved overlapping slices (1.5 mm in-plane resolution, 3 mm thickness, 1.5 mm overlap) were used to acquire a single volume, with parameters $T_R = 3800$ ms, $T_E = 90$ ms. This data underwent denoising, outlier removal, motion correction, and it

[1] http://www.developingconnectome.org/.

was subsequently super-resolved to a 1.5 mm isotropic voxel resolution [4]. All resulting images were checked for abnormalities by a paediatric neuroradiologist, and infants with major congenital malformations were excluded.

Image Selection. For this study, we use a total of 363 T_2w and FA maps of neonates born between 23–42 weeks gestational age (GA) and scanned at term-equivalent age (37–45 weeks GA). The age distribution in our dataset is found in Fig. 1.

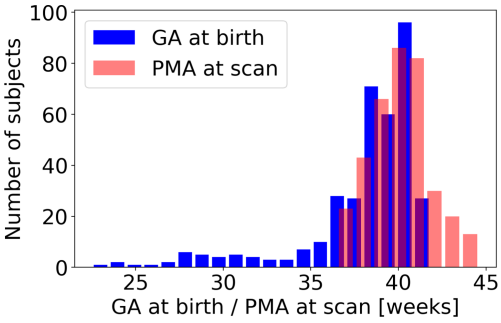

Fig. 1. Distribution of gestational ages at birth (GA) and post-menstrual ages at scan (PMA) in our imaging dataset.

Image Preprocessing. In order to use both the T_2w and FA axial slices in our registration networks, we first resampled both types of modalities into a template space of 1 mm isotropic resolution. We then affinely pre-registered both to a common 36 weeks gestational age atlas space [17] using the MIRTK software toolbox [14] and obtained the FA maps using the MRtrix3 toolbox [16]. Finally, we performed skull-stripping using the available dHCP brain masks [4], and we cropped the resulting images to a 128×128 size. Out of the 363 subjects in our dataset, we used 290 for training, 37 for validation and 36 for test, divided such that the GA at birth and post-menstrual age (PMA) at scan were kept as similar to the original distributions as possible. The validation set was used to inform us about our models' performance during training. All of our results are reported on the test set.

Network Architecture. In this study, we employ a CVAE [10] to model the registration probabilistically as proposed by [11]. Figure 2 shows the architecture at both train and inference time. In short, a pair of 2D MRI axial slices ($\mathbf{M}$ and $\mathbf{F}$) are passed through the network to learn a velocity field v, while the *exponentiation layers* (with 4 *scaling-and-squaring* [2] steps) transform it into a topology-preserving deformation field ϕ. A *Spatial Transformer* layer [6] is then used to warp (linearly resample) the moving images $\mathbf{M}$ and obtain the moved image $\mathbf{M}(\phi)$.

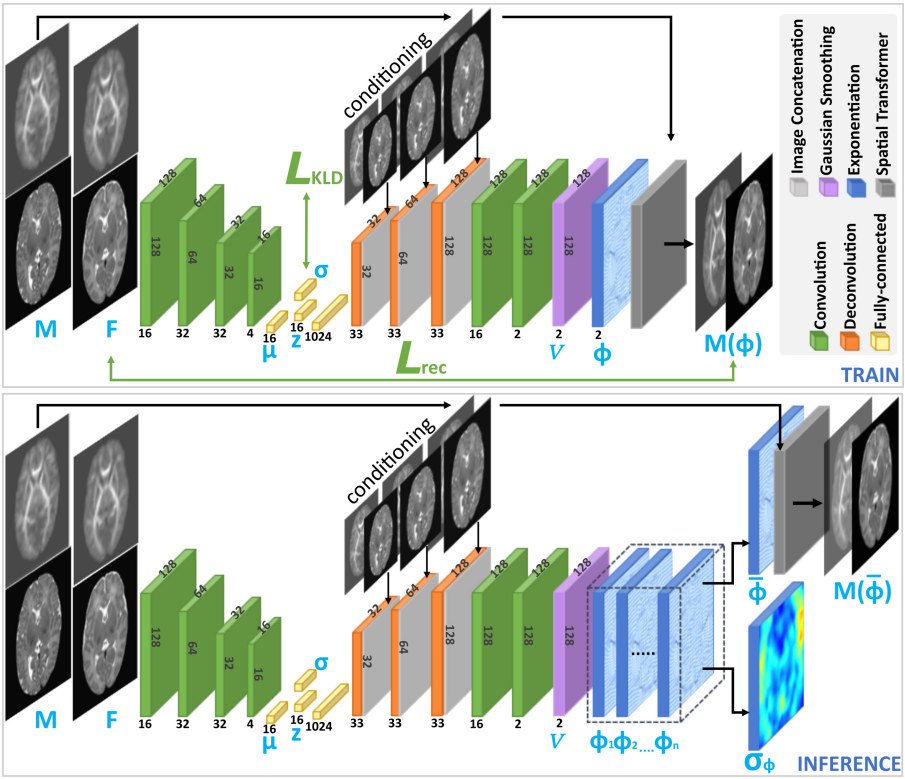

Fig. 2. We use a convolutional neural network based on the architecture proposed by [11]. During inference, we use the trained network to generate n dense displacement fields ϕ_i and create a mean displacement field $\overline{\phi}$, and its associated uncertainty map σ_ϕ.

The encoder branch is made up of four 2D convolutional layers of $16, 32, 32$, and 4 filters, respectively, with a kernel size of 3^2, followed by *Leaky ReLU* ($\alpha = 0.2$) activations [18]. The bottleneck (μ, σ, z) is fully-connected, and we kept the latent code size (16) the same as in the original paper [11]. The decoder branch is composed of three 2D deconvolutional layers of 32 filters and a kernel size of 3^2 each, followed by *Leaky ReLU* ($\alpha = 0.2$) activations. The deconvolutional layers' feature maps are concatenated with the original-sized or downsampled versions of the moving input image. Two more convolutional layers (with 16 and 2 filters, respectively) are added, followed by a Gaussian smoothing layer (kernel size 21) which outputs the velocity field v.

Training. For this study, we train three separate models on different combinations of input data. The first model is trained on pairs of structural data (T_2w-only), the second model on microstructural data (FA-only), while the third model uses both modalities as input to the network (T_2w+FA). While training

the latter, the input changes to a 4-channel tensor (moving and fixed T_2w and FA), and the conditioning to a 2-channel tensor (concatenated T_2w and FA slices of the same neonate) as shown in Fig. 2.

For each input pair, the encoder q_ω (with trainable network parameters ω) outputs the mean $\mu \in \mathbb{R}^d$ and diagonal covariance $\sigma \in \mathbb{R}^d$, from which we sample the latent vector $z = \mu + \epsilon \cdot \sigma$, with $\epsilon \sim \mathcal{N}(0, I)$. The decoder network p_γ (with trainable network parameters γ) uses the z-sample to generate a displacement field ϕ which, together with the moving image $\mathbf{M}$, produces the warped image $\mathbf{M}(\phi)$. During training, the optimizer aims to minimize: 1) the Kullback-Leibler (KL) divergence $\mathcal{L}_{KLD}$ in order to reduce the gap between the prior $p(z)$ (multivariate unit Gaussian distribution $p(z) \sim \mathcal{N}(0, I)$) and the encoded distribution $q_\omega(z|\ \mathbf{F}, \mathbf{M})$, and 2) the reconstruction loss $\mathcal{L}_{rec}$ between the fixed image $\mathbf{F}$ and warped image $\mathbf{M}(\phi)$. The loss function results in:

$$\mathcal{L} = \underbrace{KL[q_\omega(z|\mathbf{F}, \mathbf{M})\,||\,p(z)]}_{\mathcal{L}_{KLD}} + \lambda\, \underbrace{\mathcal{NCC}(\mathbf{F}(\phi^{-\frac{1}{2}}), \mathbf{M}(\phi^{\frac{1}{2}}))}_{\mathcal{L}_{rec}} \tag{1}$$

where λ is a hyperparameter set to 5000 as proposed in [11], and $\mathcal{NCC}$ is the symmetric normalised cross correlation (NCC) dissimilarity measure defined as:

$$\mathcal{NCC}(\mathbf{F}(\phi^{-\frac{1}{2}}), \mathbf{M}(\phi^{\frac{1}{2}})) = -\frac{\sum_{\mathbf{x}\in\Omega}(\mathbf{F}(\phi^{-\frac{1}{2}}) - \overline{F}) \cdot (\mathbf{M}(\phi^{\frac{1}{2}}) - \overline{M})}{\sqrt{\sum_{\mathbf{x}\in\Omega}(\mathbf{F}(\phi^{-\frac{1}{2}}) - \overline{F})^2 \cdot \sum_{\mathbf{x}\in\Omega}(\mathbf{M}(\phi^{\frac{1}{2}}) - \overline{M})^2}}$$

where $\overline{F}$ is the mean voxel value in the warped fixed image $\mathbf{F}(\phi^{-\frac{1}{2}})$ and $\overline{\mathbf{M}}$ is the mean voxel value in the warped moving image $\mathbf{M}(\phi^{\frac{1}{2}})$.

We train our models for 2500 epochs each, using the Adam optimizer with its default parameters (β_1=.9 and β_2=.999), a constant learning rate of $5 \cdot 10^{-4}$, and a L_2 weight decay factor of 10^{-5}. All networks were implemented in PyTorch.

Uncertainty-Aware Image Registration. To investigate uncertainty-aware image registration, we use our trained models to generate uncertainty maps. We achieve this at inference time by using the trained decoders to generate multiple displacement fields ϕ_i, as shown in Figs. 2 and 3. More specifically, for each subject in our test dataset, we first use the trained encoders to yield the μ and σ outputs. Then, we generate n latent vector $z = \mu + \epsilon \cdot \sigma$ samples and pass them through the trained decoder networks to generate n dense displacement fields ϕ_i. Throughout this work we set $n = 50$. From these, we obtain a mean displacement field $\overline{\phi}$ and a standard deviation displacement field σ_ϕ for each model.

For the uncertainty-aware image registration task, we combine the T_2w-only and the FA-only models into a single model, which we call T_2w+FA+uncert. This is achieved in a three-step process. First, we use the trained T_2w-only and FA-only models to generate n dense displacement fields ϕ_i, and create the modality-specific mean displacement fields $\overline{\phi}_{T2w}$ and $\overline{\phi}_{FA}$, and uncertainty maps $\sigma_{\phi_{T2w}}$ and $\sigma_{\phi_{FA}}$. Second, we calculate certainty maps ($\alpha_{\phi_{T2w}}$, $\alpha_{\phi_{FA}}$) using the

following equations:

$$\alpha_{\phi_{T2w}} = \frac{1/\sigma_{\phi_{T2w}}}{1/\sigma_{\phi_{T2w}} + 1/\sigma_{\phi_{FA}}} \;\; ; \;\; \alpha_{\phi_{FA}} = \frac{1/\sigma_{\phi_{FA}}}{1/\sigma_{\phi_{T2w}} + 1/\sigma_{\phi_{FA}}} \tag{2}$$

Finally, the T_2w+FA+uncert model's displacement field is constructed by locally weighting the contributions from each modality with the 2D certainty maps: $\phi_{T2w+FA+uncert} = \alpha_{\phi_{T2w}} \odot \overline{\phi}_{T2w} + \alpha_{\phi_{FA}} \odot \overline{\phi}_{FA}$, where $\odot$ represents element-wise multiplication. A visual explanation of these steps can be seen in Fig. 3.

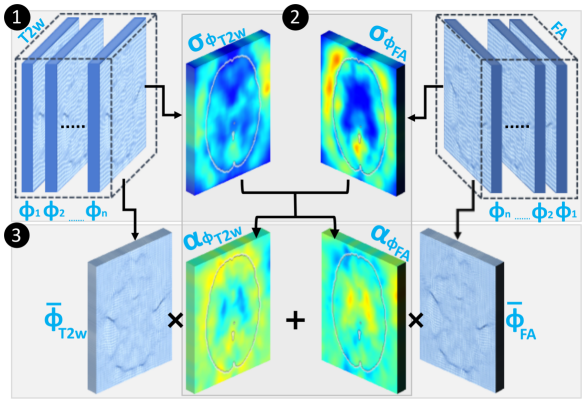

Fig. 3. The construction of certainty maps: 1) Create modality-specific mean displacement fields $\overline{\phi}_{T2w}$ and $\overline{\phi}_{FA}$, and uncertainty maps $\sigma_{\phi_{T2w}}$ and $\sigma_{\phi_{FA}}$. 2) Create modality specific certainty maps $\alpha_{\phi_{T2w}}$ and $\alpha_{\phi_{FA}}$ using Eq. 2. 3) Create the $\phi_{T2w+FA+uncert}$ displacement field by locally weighting the contributions from each modality with the 2D certainty maps.

3 Results

To validate whether our proposed model (T_2w+FA+uncert) can outperform the other three models (T_2w-only, FA-only, T_2w+FA), we carry out a quantitative evaluation on our test dataset. Each subject and template had the following tissue label segmentations obtained using the Draw-EM pipeline [13]: cerebrospinal fluid (CSF), cGM, WM, ventricles, deep gray matter (dGM), and a WM structure called the internal capsule (IC). These labels were propagated from each subject into the template space using the predicted deformation fields. The resulting Dice scores and average Hausdorff distances calculated between the warped labels and the template labels are summarised in Fig. 4, where the initial pre-alignment is shown in pink, the T_2w-only results are shown in green, the FA-only results in light blue, the T_2w+FA model in magenta, and our proposed model's results in purple (T_2w+FA+uncert). The yellow diamond points to the best performing model for each tissue type and metric.

In terms of Dice scores, our proposed model performs similarly well to the T_2w-only model in the cGM, WM, ventricles and the dGM structures, and outperforms the T_2w+FA model in the cGM, and WM labels. At the same time, the T_2w+FA+uncert obtains the lowest average Hausdorff distances in the cGM and WM regions, again outperforming the T_2w+FA model. The IC is best aligned using the FA-only model, where the T_2w-only performs worse, while the models which incorporate FA maps perform comparably well.

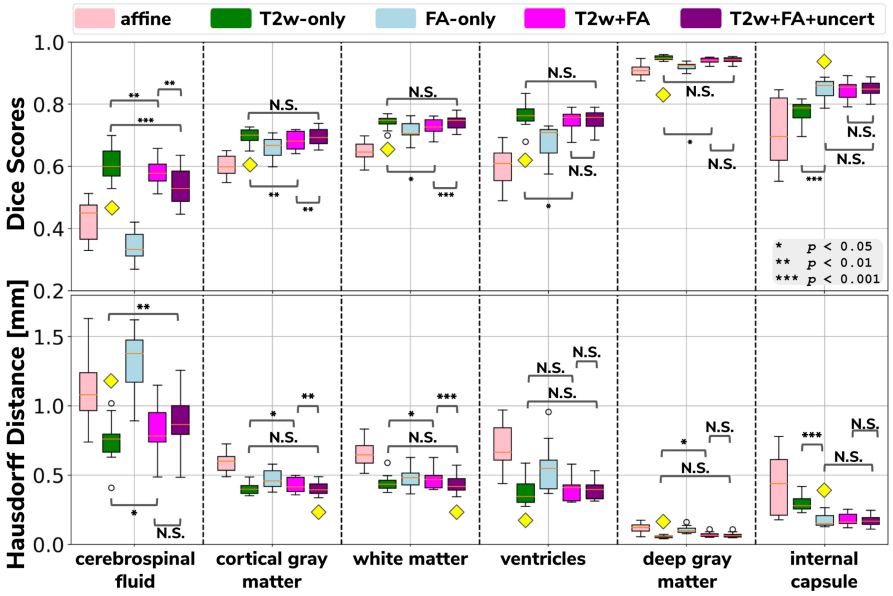

Fig. 4. The results on our test dataset for all four methods, together with the initial affine alignment. The yellow diamond highlights the model which performed best for its respective tissue type and metric. (***N.S.*** *means "not-significant"*) (Color figure online)

Next, we used data from 10 neonatal subjects scanned around 40 weeks PMA to produce average uncertainty maps for our three trained models. For each subject $j \in [1, 10]$ and model $m \in \{T_2$w-only, FA-only, T_2w+FA$\}$, we obtained an uncertainty map $\sigma^j_{\phi_m}$, and averaged them across the subjects. Figure 5 shows these average uncertainty maps (in the template space), overlaid on top of the fixed images which were used for training. By combining the uncertainty maps from the trained T_2w-only and FA-only models, we can obtain modality-dependent certainty maps (see Eq. 2) which are shown on the last row.

The T_2w-only model (first row in Fig. 5) yields high uncertainty in dGM regions (cyan arrow) where there is little contrast, as well as in difficult brain areas, such as the cGM folds. Similarly, the FA-only model (first row in Fig. 5) shows high uncertainty in low contrast cortical areas (yellow arrow) and low uncertainty in the high contrast WM structures such as the IC region (cyan

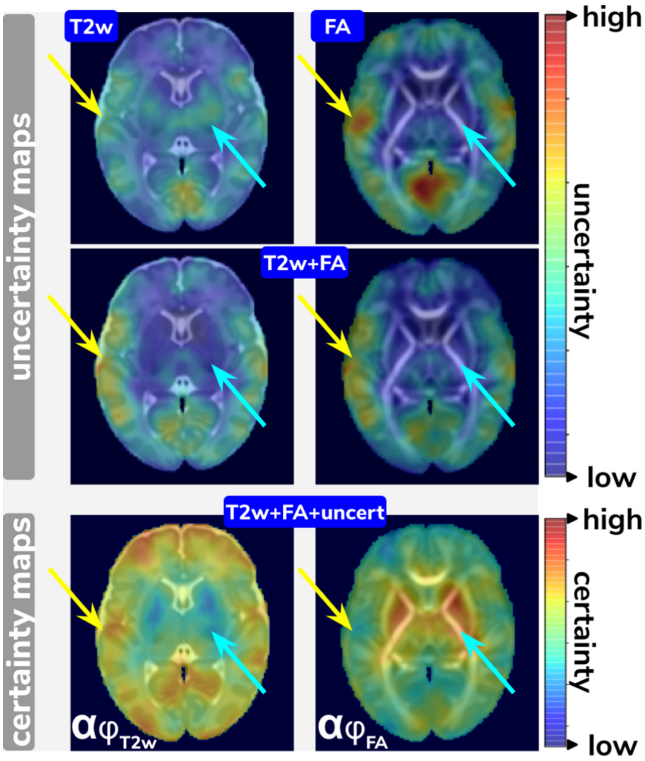

Fig. 5. Average uncertainty maps for the T_2w- and FA-only (on the first row) and the T_2w+FA models (on the second row). The cyan arrows point to the same region of dGM, while the yellow arrows point to the same region of cGM for all models and modalities, respectively. The last row shows the modality-dependent certainty maps obtained from the T_2w-only and FA-only models. (Color figure online)

arrow). When using both modalities (second row in Fig. 5), the uncertainty becomes low in the dGM regions, as the model is being helped by the extra FA channel, but becomes higher in the cGM regions (yellow arrows) when compared to the T_2w-only model. This could be caused by the high uncertainty in the low contrast cortical regions of the FA channel. The average certainty maps are shown on the last row, where we can see that in our proposed model the combined displacement field will depend more on the FA-only model in the WM tracts regions as seen in the $\alpha_{\phi_{FA}}$ certainty map (cyan arrow), and on the T_2w-only model for the cortical regions (yellow arrows).

4 Discussion and Future Work

This paper presents a novel solution for multi-channel registration, which combines FA and T_2w data driven displacement fields based on their respective

uncertainty maps. The quantitative evaluation performed on 36 neonatal subjects from the dHCP project showed that the proposed certainty based fusion of structural and microstructural channels improves overall alignment when compared to models trained on either single-channel or multi-channel data. The main limitations of this work are: the use of a single latent code size and smoothing kernel, the use of 2-D axial slices only, and no comparison with other probabilistic registration frameworks (*e.g.,* [6]). Future work will focus on further improving the registration accuracy in the cortical regions, on adapting our work to 3-D datasets, and on exploring the aforementioned limitations.

Acknowledgments. 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. 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)
3. Barnett, M.L., et al.: Exploring the multiple-hit hypothesis of preterm white matter damage using diffusion MRI. NeuroImage: Clin. **17**, 596–606 (2018)
4. Christiaens, D., et al.: Scattered slice SHARD reconstruction for motion correction in multi-shell diffusion MRI. NeuroImage **225**, 117437 (2021). https://doi.org/10.1016/j.neuroimage.2020.117437
5. Cordero-Grande, L., Hughes, E.J., Hutter, J., Price, A.N., Hajnal, J.V.: Three-dimensional motion corrected sensitivity encoding reconstruction for multi-shot multi-slice MRI: Application to neonatal brain imaging. Magn. Reson. Med. **79**(3), 1365–1376 (2018)
6. 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
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. Hughes, E.J., et al.: A dedicated neonatal brain imaging system. Magn. Reson. Med. **78**(2), 794–804 (2017)
9. Hutter, J., et al.: Time-efficient and flexible design of optimized multishell HARDI diffusion. Magn. Reson. Med. **79**(3), 1276–1292 (2018)
10. Kingma, D.P., Rezende, D.J., Mohamed, S., Welling, M.: Semi-supervised learning with deep generative models (2014). https://arxiv.org/abs/1406.5298
11. 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
12. Kuklisova-Murgasova, M., Quaghebeur, G., Rutherford, M.A., Hajnal, J.V., Schnabel, J.A.: Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. Med. Image Anal. **16**, 1550–1564 (2012)
13. Makropoulos, A., Gousias, I.S., Ledig, C., Aljabar, P., Serag, A., Hajnal, J.V., Edwards, A.D., Counsell, S.J., Rueckert, D.: Automatic whole brain MRI segmentation of the developing neonatal brain. IEEE Trans. Med. Imaging **33**(9), 1818–1831 (2014)
14. 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)
15. Schuh, A., et al.: Unbiased construction of a temporally consistent morphological atlas of neonatal brain development. bioRxiv (2018)
16. Tournier, J.D., et al.: MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. NeuroImage **202**, 116137 (2019)
17. Uus, A., et al.: Multi-channel 4D parametrized atlas of macro- and microstructural neonatal brain development. Front. Neurosci. **15**, 721 (2021). https://doi.org/10.3389/fnins.2021.661704
18. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network (2015). https://arxiv.org/abs/1505.00853