

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Deep learning analysis of multi-modal neonatal MRI

Grigorescu, Irina

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



DEPARTMENT OF BIOMEDICAL ENGINEERING

Deep learning analysis of multi-modal neonatal MRI

Author:
Irina Grigorescu

Supervisors:
Dr Maria Deprez
Dr Marc Modat

A thesis submitted for the degree of

Doctor of Philosophy

July 19, 2023

pentru părinții mei

Declaration

I, Irina Grigorescu, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Irina Grigorescu
July 19, 2023

Acknowledgements

First and foremost, I would like to thank my supervisors for their guidance, advice and help over the last few years. In particular, I would like to thank Maria Deprez for being a fantastic supervisor, with a brilliant mind, who has always had an open door for me, and who has always been there to guide and support me through thick and thin. I am incredibly grateful to have been Maria's PhD student, as she has shown me the perfect mixture of kindness and understanding, while also pushing me to excel.

I would also like to thank Marc Modat, without whom this journey would not have started. Thank you, Marc, for believing in me, guiding and supporting me throughout the years.

In terms of the finishing touches of this thesis, I would like to express my gratitude to Dr Yipeng Hu and Dr Ivana Išgum. Upon incorporating your valuable feedback and suggestions, numerous sections of my manuscript have gained enhanced clarity and coherence.

I would like to thank everyone who has been involved in my project. Many thanks to Dafnis, Andy and Donald for providing me with extra insight and support through the TPC meetings. To Lucy, thank you for sharing your thoughts and expertise with me. I would also like to extend my gratitude to Dr Alena Uus, whose help, insights and guidance have been invaluable to my journey. To Irme, Marta, David and Samuel, thank you for always taking the time to meet and providing me with your knowledge and support.

I want to thank all of my friends who have supported me during this PhD. Thank you Irme, Maria, Nooshin, Marta, and the entire CoolKids and FNIAG teams for your friendship. I would also like to thank all of my friends from back home: Raluca, Andru, Claudiu, Ștefan, Elena and Noah. Your friendship means the world to me. Thank you, Eileen Rose and Roxana Agafiței, for being the moral support I needed in the last year of my PhD.

Finally, I would like to thank my family. First, my UK family. Thank you, Danny, for always being there for me, for always believing in me and for always pushing me to be my best self. Mel, thank you for being my sister away from home, for all the teas and biscuits we had together and all the TV shows that made us laugh. Raj, Dave and Manoon, thank you for being my family when I needed one most.


Mami și tati, vă mulțumesc că m-ați crescut și că m-ați susținut în toți anii aceștia în care am fost departe de casă. Fără dragostea și sacrificiile voastre nu aș fi ajuns așa departe. Răzvan, îți mulțumesc că mi-ai fost mereu aproape și că m-ai ajutat cu tot ce am avut nevoie. În plus, fără tine și fără tata nu aș fi crescut iubind tot ce înseamnă știință. Mami, îți mulțumesc; tu mi-ai spus mereu să am curaj și încredere în mine, m-ai învățat să îmi înfrunt fricile, și să mă ridic atunci când am căzut. Dana, Andreea și Matei, vă mulțumesc că ați fost alături de mine și că m-ați susținut în toți acești ani. Fără ajutorul vostru nu aș fi reușit să devin persoana de astăzi.






Publication List

Peer-reviewed Journal Papers

First author:


- (J1) **Grigorescu I**, Vanes L, Uus AU, Batalle D, Cordero-Grande L, Nosarti C, Edwards AD, Hajnal JV, Modat M and Deprez M, “*Harmonized Segmentation of Neonatal Brain MRI*” in *Front. Neurosci.* (2021)
 doi.org/10.3389/fnins.2021.662005

Co-author:



- (J2) Uus AU, **Grigorescu I**, van Poppel MPM, Steinweg JK, Roberts TA, Rutherford MA, Hajnal JV, Lloyd DFA, Pushparajah K and Deprez M, “*Automated 3D reconstruction of the fetal thorax in the standard atlas space from motion-corrupted MRI stacks for 21-36 weeks GA range*” in *Medical Image Analysis* (2022)
 doi.org/10.1016/j.media.2022.102484
- (J3) Taoudi-Benchekroun Y, Christiaens D, **Grigorescu I**, Gale-Grant O, Schuh A, Pietsch M, Chew A, Harper N, Falconer S, Poppe T, Hughes E, Hutter J, Price AN, Tournier J-D, Cordero-Grande L, Counsell SJ, Rueckert D, Arichi T, Hajnal JV, Edwards AD, Deprez M and Batalle D, “*Predicting age and clinical risk from the neonatal connectome*” in *NeuroImage* (2022)
 doi.org/10.1016/j.neuroimage.2022.119319
- (J4) Uus AU, **Grigorescu I**, Pietsch M, Batalle D, Christiaens D, Hughes E, Hutter J, Cordero-Grande L, Price AN, Tournier J-D, Rutherford MA, Counsell SJ, Hajnal JV, Edwards AD and Deprez M, “*Multi-Channel 4D Parametrized Atlas of Macro- and Microstructural Neonatal Brain Development*” in *Front. Neurosci.* (2021)
 doi.org/10.3389/fnins.2021.661704



Peer-reviewed Conference Papers

First author:

- (C1) **Grigorescu I**, Uus AU, Christiaens D, Cordero-Grande L, Hutter J, Batalle D, Edwards AD, Hajnal JV, Modat M and Deprez M, *Attention-Driven Multi-channel Deformable Registration of Structural and Microstructural Neonatal Data* in PIPPI, Lecture Notes in Computer Science, Springer (2022)
 doi.org/10.1007/978-3-031-17117-8_7
- (C2) **Grigorescu I**, Uus AU, Christiaens D, Cordero-Grande L, Hutter J, Batalle D, Edwards AD, Hajnal JV, Modat M and Deprez M, *Uncertainty-Aware Deep Learning Based Deformable Registration* in UNSURE, Lecture Notes in Computer Science, Springer (2021)
 doi.org/10.1007/978-3-030-87735-4_6
- (C3) **Grigorescu I**, Cordero-Grande L, Batalle D, Edwards AD, Hajnal JV, Modat M and Deprez M, “*Harmonised Segmentation of Neonatal Brain MRI: A Domain Adaptation Approach*” in PIPPI, Lecture Notes in Computer Science, Springer (2020)
 doi.org/10.1007/978-3-030-60334-2_25
- (C4) **Grigorescu I**, Uus AU, Christiaens D, Cordero-Grande L, Hutter J, Edwards AD, Hajnal JV, Modat M and Deprez M, “*Diffusion Tensor Driven Image Registration: A Deep Learning Approach*” in WBIR, Lecture Notes in Computer Science, Springer (2020)
 doi.org/10.1007/978-3-030-50120-4_13
- (C5) **Grigorescu I**, Cordero-Grande L, Edwards AD, Hajnal JV, Modat M and Deprez M, “*Investigating Image Registration Impact on Preterm Birth Classification: An Interpretable Deep Learning Approach*” in PIPPI, Lecture Notes in Computer Science, Springer (2019)
 doi.org/10.1007/978-3-030-32875-7_12



Co-author:

- (C6) Uus AU, Ayub M-U, Gartner A, Kyriakopoulou V, Pietsch M, **Grigorescu I**, Christiaens D, Hutter J, Cordero-Grande L, Price A, Batalle D, Counsell S, Hajnal JV, Edwards AD, Rutherford MA and Deprez M, “*Segmentation of Periventricular White Matter in Neonatal Brain MRI: Analysis of Brain Maturation in Term and Preterm Cohorts*” in PIPPI, Lecture Notes in Computer Science, Springer (2022)
 doi.org/10.1007/978-3-031-17117-8_9
- (C7) Ramirez Gilliland P, Uus AU, van Poppel MPM, **Grigorescu I**, Steinweg JK, Lloyd DFA, Pushparajah K, King A and Deprez M *Automated Multi-class Fetal Cardiac Vessel Segmentation in Aortic Arch Anomalies Using T2-weighted 3D Fetal MRI* in PIPPI, Lecture Notes in Computer Science, Springer (2022)
 doi.org/10.1007/978-3-031-17117-8_8






-
- (C8) Uus AU, Pietsch M, **Grigorescu I**, Christiaens D, Tournier J-D, Cordero-Grande L, Hutter J, Edwards AD, Hajnal JV and Deprez M, “*Multi-channel Registration for Diffusion MRI: Longitudinal Analysis for the Neonatal Brain*” in WBIR, Lecture Notes in Computer Science, Springer (2020)
 doi.org/10.1007/978-3-030-50120-4_11
- (C9) Lourenço A, Kerfoot E, **Grigorescu I**, Scannell CM, Varela M and Correia TM, “*Automatic Myocardial Disease Prediction from Delayed-Enhancement Cardiac MRI and Clinical Information*” in STACOM, Lecture Notes in Computer Science, Springer (2020)
 doi.org/10.1007/978-3-030-68107-4_34

Peer-reviewed Conference Abstracts

First author:

- (A1) **Grigorescu I**, Cordero-Grande L, Edwards AD, Hajnal JV, Modat M and Deprez M, “*Interpretable Convolutional Neural Networks for Preterm Birth Classification*” MIDL Extended Abstract Track (2019)
 [10.48550/arxiv.1910.00071](https://doi.org/10.48550/arxiv.1910.00071)  openreview.net/forum?id=SyevkEaEcE

Co-author:

- (A2) Uus AU, van Poppel MPM, Steinweg JK, **Grigorescu I**, Collado AE, Ramirez Gilliland P, Roberts TA, Hajnal JV, Rutherford M, Lloyd DFA, Pushparajah K and Deprez M, “*3D MRI atlases of congenital aortic arch anomalies and normal fetal heart: application to automated multi-label segmentation*” in ISMRM (2022)
 archive.ismrm.org/2022/0270.html  doi.org/10.1101/2022.01.16.476503
- (A3) Neves Silva S, **Grigorescu I**, Uus AU, Steinweg J, Deprez M, Hajnal J, Pushparajah K, Hutter J and De Vita E, “*Evaluation of dynamic 2D-EPI acquisitions for fetal brain tracking with neural networks*” in ISMRM (2022)
 archive.ismrm.org/2022/1820.html
- (A4) Singh M, Neves Silva S, Uus AU, **Grigorescu I**, Rutherford M, Hajnal J and Hutter J, “*Real-time survey base estimation of gestational age to guide a fetal MRI scan*” in ISMRM (2022)
 archive.ismrm.org/2022/3912.html
- (A5) Uus AU, **Grigorescu I**, van Poppel MPM, Hughes E, Steinweg J, Roberts TA, Lloyd DFA, Pushparajah K and Deprez M, “*3D UNet with GAN Discriminator for Robust Localisation of the Fetal Brain and Trunk in MRI with Partial Coverage of the Fetal Body*” in Med NeurIPS (2020)
 [medneurips2020/51_med_neurips_2020-final-1811.pdf](https://arxiv.org/abs/2005.11811)

Abstract

Analysis of magnetic resonance images (MRI) of the neonatal brain comes with unique challenges. The rapid development results in changes in both shape and appearance of the neonatal brain scanned at different post-menstrual weeks. These changes affect outputs of image analysis tools, such as image registration or segmentation, making interpretation of the results difficult.

The aim of this PhD project is to develop deep learning image segmentation and registration tools to address challenges in analysing the developing neonatal brain MRI. While accurate segmentation of neonatal brain MRI has been achieved by existing classical segmentation techniques, these are sensitive to MRI acquisition protocols, making volumetric comparisons between subjects from different studies unreliable. I therefore propose harmonised deep learning-based segmentation for neonatal MRI. At the same time, traditional medical image registration methods can be misguided by the rapid MRI contrast changes due to ongoing brain tissue maturation in the first weeks of life. To alleviate this problem, I propose a multi-channel attention based deep learning registration approach that selects the most salient features from multiple image modalities to improve alignment of individual MR images to a common atlas space.

As a prerequisite for the contributions, the first chapter introduces the neonatal brain, and describes the main MRI modalities, as well as two neonatal datasets, which were utilised in this thesis. The second and third chapters lay the groundwork for the methods used throughout this thesis, with a focus on classical and deep learning image registration and segmentation algorithms. A survey of state-of-the-art deep learning based medical image registration and segmentation techniques follows, with the aim of presenting some of the baseline models used throughout this thesis, as well as more advanced techniques, such as unsupervised domain adaptation and visual attention.

The three novel chapters of the thesis describe my contributions. First, I investigated deep learning domain adaptation algorithms to suppress the domain shift between a target and a source dataset, thus making it feasible to predict on unseen data distributions. My proposed image-space domain adaptation model combined with data augmentation achieved the best solution for harmonising tissue segmentation maps of two neonatal datasets. I have shown that there were no significant differences in tissue volumes and cortical thickness measures derived from the harmonised segmentations on a subset of the datasets matched for gestational age at

birth and postmenstrual age at scan. Second, I developed a novel attention-based deep learning multi-channel registration model that learns spatially varying attention maps needed to fuse different modalities, thus taking advantage of their complementary nature. I applied the technique to align multi-channel datasets composed of structural T_2 -weighted (T_2w) MRI and fractional anisotropy (FA) maps derived from diffusion MRI to the atlas space. 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. Finally, I expanded the registration model from the previous chapter to align multi-channel data composed from structural T_2w MRI and diffusion tensor maps into atlas space, which further improved alignment of white matter tracts.

In my PhD thesis I proposed solutions to tackle some of the challenges in analysis of the neonatal MRI, when the developing brain changes both shape and MRI tissue contrast as it grows. The techniques will support accurate image segmentation independent of the acquisition protocol and multi-channel registration to atlas space, that can take advantage of different information content of various MRI modalities. These techniques will help to improve reliability and interpretability of downstream neuroimaging analyses.

Contents

Abstract	8
1 Introduction	16
1.1 The neonatal brain	16
1.2 Imaging of the developing brain	18
1.2.1 Structural MRI	19
1.2.2 Microstructural MRI	21
1.2.3 Acquisition and preprocessing of neonatal MRI	24
1.3 Challenges in analysis of the neonatal brain MRI	28
1.4 Thesis contributions	30
1.5 Thesis outline	31
2 Medical image registration and segmentation for neonatal brain MRI	33
2.1 Medical image registration	34
2.1.1 Image similarity measures	37
2.1.2 Image registration regularisation through penalty terms	40
2.1.3 Transformation models	41
2.1.4 Diffusion tensor image registration	47
2.2 Medical image segmentation	49
2.2.1 Registration-based approaches	50
2.2.2 Intensity-based approaches	53
2.2.3 Medical image segmentation frameworks	59
2.2.4 Medical image segmentation evaluation	61
3 Deep learning for medical image analysis	65
3.1 Deep learning theory	66
3.1.1 Artificial neural networks	66
3.1.2 Training a neural network	69
3.1.3 Network architectures	76
3.1.4 Network training strategies	81
3.2 Deep learning for medical image analysis	84
3.2.1 Deep learning-based medical image registration	84
3.2.2 Deep learning-based medical image segmentation	92
3.2.3 Visual attention	100
3.2.4 Domain adaptation	106

4	Harmonised segmentation of neonatal brain MRI	112
4.1	Introduction	113
4.2	Methods	114
4.2.1	Data acquisition and preprocessing	114
4.2.2	Unsupervised domain adaptation models	116
4.2.3	Network architectures	116
4.2.4	Training	118
4.3	Results	120
4.3.1	dHCP test dataset	120
4.3.2	Validation of data harmonisation	122
4.3.3	Analysis of harmonised cortical substructures	129
4.4	Discussion and future work	132
5	Attention-driven multi-channel deformable registration of structural and microstructural neonatal data	136
5.1	Introduction	137
5.2	Methods	138
5.2.1	Data acquisition and preprocessing	138
5.2.2	Network architectures	138
5.2.3	Training the networks	142
5.3	Results	145
5.3.1	Quantitative evaluation	145
5.3.2	Qualitative results	150
5.4	Discussion and future work	154
6	Diffusion tensor driven deep learning image registration	156
6.1	Introduction	157
6.2	Methods	157
6.2.1	Data acquisition and preprocessing	157
6.2.2	Network architectures	158
6.2.3	Tensor reorientation	159
6.2.4	Training the networks	160
6.3	Results	163
6.3.1	Quantitative evaluation	163
6.3.2	Qualitative results	169
6.4	Discussion and future work	173
7	Conclusions	175
7.1	Limitations and future work	177
7.1.1	Inclusion of multiple imaging modalities and labels	177
7.1.2	Further exploring the image synthesis avenue	177
7.1.3	Identifying abnormal developmental patterns	178
Appendices		179
A	Grouping of cortical substructures 1/2	179
B	Grouping of cortical substructures 2/2	180
Bibliography		217

List of Figures

1.1	Brain development timeline	17
1.2	Comparison of T_2w MRI mid-brain slices of a term and very preterm neonates	18
1.3	Axial, coronal and sagittal T_1w and T_2w MRI slices of a neonate, an infant, and a young adult	20
1.4	Axial slices of five neonates showing both the color-coded directionality map and their respective T_2w images	21
1.5	Example of diffusion trajectory, ellipsoids and tensors for isotropic unrestricted, isotropic restricted and anisotropic restricted diffusion	22
1.6	Schematic representation of the diffusion tensor ellipsoid	23
1.7	Schematic representation of the relationship between the diffusion tensor ellipsoid, its eigenvalues, eigenvectors and its trace	24
1.8	Example of ePrime and dHCP neonates	26
1.9	Density plot of variance of the Laplacian computed on the T_2w MRI images of both ePrime and dHCP neonatal datasets	27
1.10	Axial slices of three neonatal atlases at 38 weeks, 41 weeks, and 44 weeks, respectively	29
2.1	Example of a spatial mapping φ between fixed and moving images	35
2.2	2D global transformations	36
2.3	Illustration of image resampling in the context of image registration	37
2.4	Example of different 1D interpolation strategies	37
2.5	Illustration of the Demons algorithm	43
2.6	DTI reorientation	48
2.7	Illustration of medical image segmentation based on a single atlas registration approach	50
2.8	Illustration of medical image segmentation based on the multi atlas registration approach	51
2.9	Illustration of medical image segmentation based on the probabilistic atlas registration approach.	52
2.10	K-nearest neighbours example	55
2.11	K-Means clustering example	56
2.12	Expectation-Maximization example	58
2.13	Draw-EM pipeline	60

2.14	Schematic example of a ground truth binary segmentation and a predicted segmentation, together with the four basic measures: TP, TN, FP, FN.	62
2.15	Confusion matrix showing the TP, TN, FP and FN measures	62
2.16	Schematic representation of the directed Hausdorff distance between two sets of points X and Y	64
3.1	Artificial neural networks	67
3.2	Example of convolutional neural network layers	68
3.3	Example of convolutional filter behaviour	69
3.4	Example of pooling layer behaviour	69
3.5	Example of activation functions	70
3.6	Softmax activation function	71
3.7	Normalisation layers	72
3.8	Example of optimizers	74
3.9	Example of learning rate schedulers	75
3.10	Schematic example of an autoencoder and a variational autoencoder	77
3.11	U-Net architectures for both 2D and 3D applications	79
3.12	The nnU-Net pipeline	80
3.13	Vanilla GAN	82
3.14	Image-to-image translation (pix2pix)	83
3.15	Cycle-GAN architecture	84
3.16	Deep similarity based registration methods	85
3.17	Fully supervised deep learning registration methods	86
3.18	Dual supervised deep learning registration methods	87
3.19	Weakly supervised deep learning for multi-modal deformable registration	88
3.20	Unsupervised deep learning registration methods	89
3.21	Voxelmorph unsupervised image registration framework	89
3.22	Diffeomorphic Voxelmorph unsupervised image registration framework	90
3.23	Diffeomorphic unsupervised probabilistic registration network framework	91
3.24	The cascade model	93
3.25	Architecture of HighResNet	94
3.26	DeepMedic segmentation model	95
3.27	Residual connection	96
3.28	Dense connection	96
3.29	Inception module	97
3.30	Binary cross entropy	99
3.31	Squeeze-and-excitation block	102
3.32	Channel attention block	103
3.33	Spatial attention block	103
3.34	Mixed channel and spatial attention block	104
3.35	Non-local attention	105
3.36	Domain adaptation	106
3.37	Intensity distribution of MRI axial slices from the dHCP and ePrime datasets	107
3.38	Adversarial domain adaptation in the latent space	109

3.39	Adversarial domain adaptation in the image space	110
4.1	Age distribution of subjects in our dHCP and ePrime datasets	115
4.2	The baseline image segmentation model	117
4.3	The latent space domain adaptation model	117
4.4	The image space domain adaptation model	118
4.5	Quantitative results of our six models on the dHCP test dataset	121
4.6	Comparison of volume measures for our 6 tissue types	124
4.7	The association between PMA and mean cortical thickness before and after applying the data harmonisation models on the matched dHCP and ePrime subsets	125
4.8	Comparison of local mean cortical thickness measures before and after data harmonization	127
4.9	Example predicted segmentation maps for the best performing models	128
4.10	Example of a neonate from the ePrime dataset with 32.86 weeks GA at birth and 39.86 weeks PMA at scan	129
4.11	Mean cortical thickness measures before and after harmonising the tissue segmentation maps	130
4.12	Comparison of cortical thickness measures for the whole cortex and for each of the 11 cortical subregions between term and preterm-born neonates	131
4.13	Language composite score against predicted left and right frontal cortical thickness measures before and after harmonising the tissue segmentation maps	131
5.1	Multi-channel image registration network during training	139
5.2	Multi-channel image registration network at inference	140
5.3	The construction of uncertainty-aware deformation fields	141
5.4	Attention-based image registration network architecture	142
5.5	Line plots showing median Dice scores for cGM and IC structures	146
5.6	Line plots showing median average surface distances for cGM and IC structures	147
5.7	Average α_{T2w} attention maps for the proposed attention multi- channel registration network	151
5.8	Average α_{T2w} and α_{FA} attention maps for the proposed attention multi-channel registration network for all values of λ_{FA}	152
5.9	Average α_{T2w} and α_{FA} attention maps for the uncertainty-aware, as well as the proposed attention models	153
6.1	Diffusion tensor driven multi-channel image registration network	159
6.2	Diffusion tensor driven multi-channel attention image registration network	160
6.3	CC scores and average OVL values of WM and IC voxels	168
6.4	Average α_{T2w} and α_{DTI} attention maps for the 2D attention multi- channel registration network	169
6.5	Average α_{T2w} and α_{DTI} attention maps for the 3D attention multi- channel registration network	170
6.6	Average DT images	172

List of Tables

4.1	Number of scans in different datasets used for training, validation and testing the models	115
4.2	Dice Scores obtained on the dHCP test set for the cortical parcellation network	122
5.1	Number of scans in different datasets used for training, validation and testing the models	138
5.2	Single- and multi-channel experiment setups used in this study, for different values of hyperparameters λ_{FA} and λ_{T2w}	145
5.3	Dice scores and average surface distances for the 3D T2w and FA experiments	149
6.1	Number of scans in different datasets used for training, validation and testing the models	158
6.2	Single- and multi-channel 2D and 3D experiment setups used in this study, for different values of hyperparameters λ_{DTI} and λ_{T2w}	162
6.3	Dice scores and average surface distances for the 2D T2w and DTI experiments	164
6.4	Dice scores and average surface distances for the 3D T2w and DTI experiments	166
1	Grouping of cortical substructures (part 1/2)	179
2	Grouping of cortical substructures (part 2/2)	180

Introduction

Medical image analysis is an important tool for understanding the anatomy and function of the human body in a non-invasive way. This holds true especially for neuroimaging studies performed using magnetic resonance imaging (MRI), such as understanding neurodevelopment, or identifying causes of neurocognitive problems. In this work, the focus is on both structural and microstructural MRI analysis of the developing brain in the neonatal period.

1.1 The neonatal brain

A normal pregnancy lasts around 40 weeks (9.2 months), with delivery between 37 and 42 weeks gestational age (GA) at birth being considered normal [1], while premature birth happens before the 37 weeks GA threshold. According to the World Health Organisation (WHO), approximately 1 in 10 babies are born prematurely every year, and it is one of the leading causes of neonatal mortality [1, 2]. In recent years, advances in perinatal care have managed to increase the survival rate of infants born before 30 weeks GA at birth, however, approximately 20% of these will suffer from behavioural or developmental disabilities, while 10% will develop motor deficits [3].

During early life, brain development is characterised by rapid changes such as myelination, cortical folding and evolving microstructure (see Figure 1.1). After the formation of the neural tube [5], which marks the beginning of development in the vertebrate nervous system, the embryonic brain differentiates into three distinct structures: the forebrain, the midbrain, and the hindbrain [4]. The forebrain will eventually give rise to the telencephalon (cerebral cortex), and the diencephalon (thalamus, hypothalamus, and other structures). This is then followed by a complex series of neurodevelopmental events which are centered around neuronal pro-

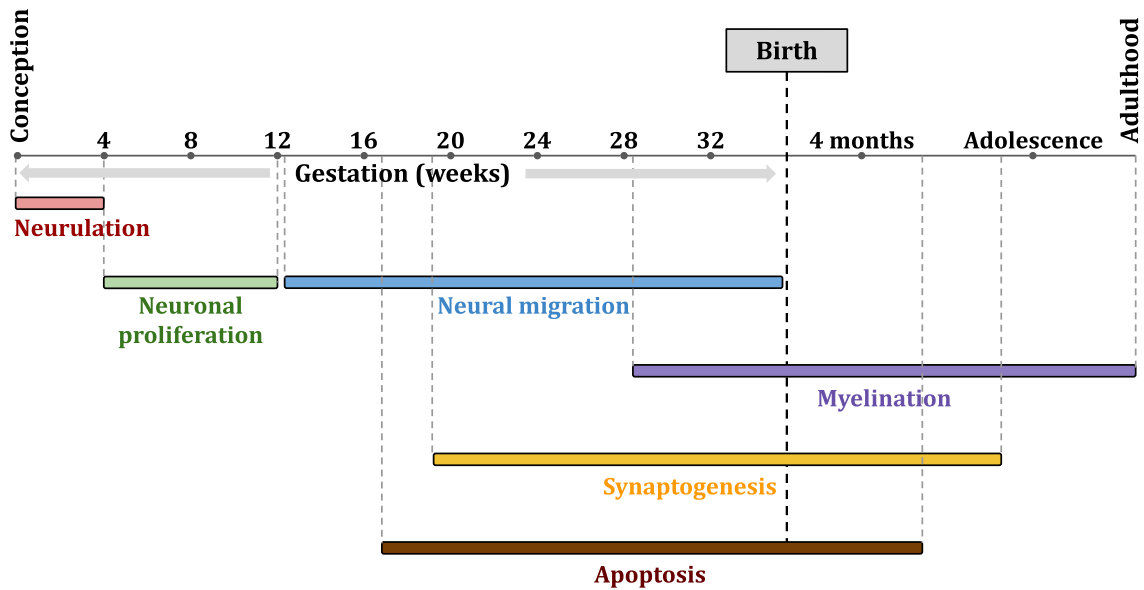


Figure 1.1: Brain development timeline. Image adapted from Tau *et al.* [4].

liferation, migration and differentiation [6]. Around 16 weeks GA, a process called apoptosis, or programmed cell death, begins as a way of eliminating unwanted cells. In fact, approximately half of the neurons created during neurogenesis are culled by the end of adolescence [4]. The formation of synapses between neurons begins around 20 weeks GA and it is an integral process in the overall architecture of brain connectivity [6]. Synaptogenesis continues throughout development and into adolescence. Around week 28 GA, the brain's neuronal axons start to myelinate, a process which involves the wrapping of neuronal axons with insulating layers made up of protein and fatty substances known as the myelin sheath, a protective covering which allows electrical impulses to transmit quickly and efficiently along the nerve cells. By 2 years of age, most brain regions show adult level myelination, with a few areas which continue to myelinate throughout adolescence and into adulthood [4].

These processes are an integral part of *in utero* brain development, and continue in early postnatal life. Preterm birth can disrupt these developmental processes, resulting in lifelong neurocognitive and neurobehavioural problems [7]. Furthermore, studies have shown that premature-born neonates have reduced cortical folding [8], as well as enlarged cerebrospinal fluid (CSF) volumes and white matter (WM) abnormalities when compared to full-term controls [9, 10]. Figure 1.2 shows example axial, coronal and sagittal slices of two neonates scanned at term-equivalent age (around 41 weeks post-menstrual age (PMA) at scan), with the first row showing a preterm-born neonate (24.7 weeks GA at birth), while the second row shows an infant born at term (41 weeks GA at birth) [11].

Magnetic resonance (MR) imaging is an excellent source for potential biomarkers of neurodevelopmental outcomes [12, 13]. However, their predictive capability has so far been limited by the rapid changes in shape and MR contrasts of the developing brain, which can easily mask effects related to preterm birth or early signs of disease [7]. This highlights the importance of understanding the differences between normal

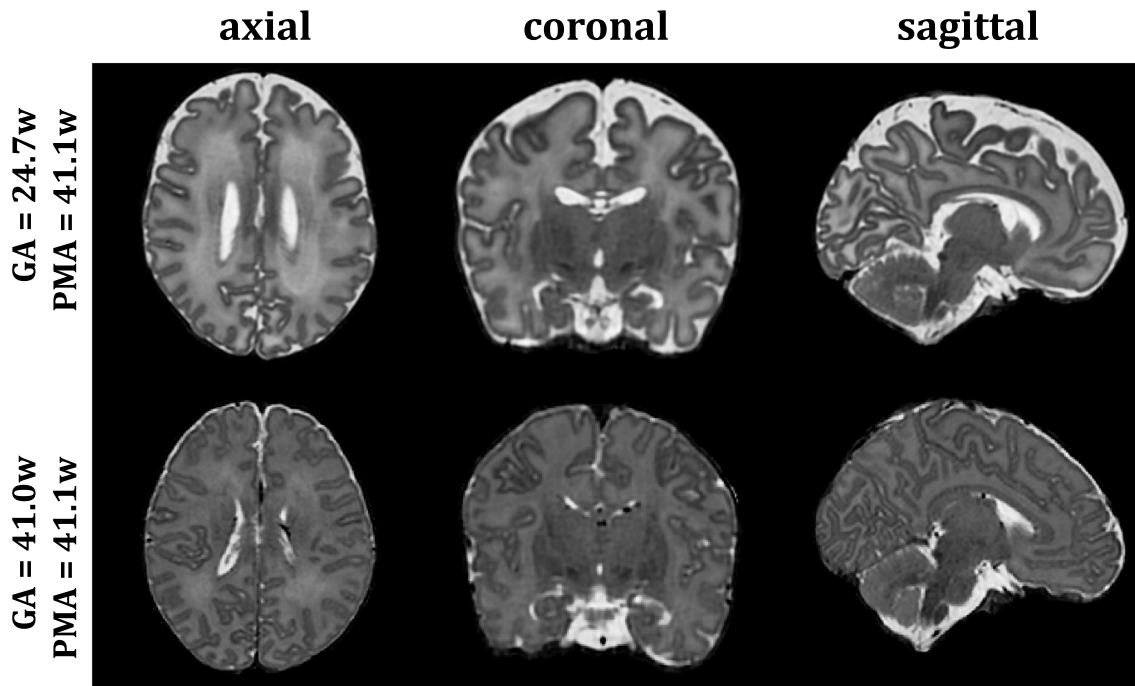


Figure 1.2: Comparison of T_2 -weighted (T_2w) MRI mid-brain slices between a term (41.0w GA at birth) and very preterm (24.7w GA at birth) neonates both scanned at 41.1w PMA [11].

and abnormal brain development during early life, which can ultimately lead to improved care for the infants and their families.

1.2 Imaging of the developing brain

MRI is a powerful and highly versatile imaging modality that uses the principles of nuclear magnetic resonance first observed by Bloch and Purcell in 1946 [14, 15]. This technique is able to produce images by spatially varying the phase and frequency of the energy being absorbed and emitted by the imaged object, a method that was proposed in 1973 in the seminal papers by Lauterbur and Mansfield [16, 17]. MRI relies on observing the way atomic nuclei respond and interact with an applied magnetic field. In clinical MRI, the focus is entirely on the hydrogen proton, the most abundant element in the human body.

In the field of biomedical sciences, MR imaging is widely used for studying anatomy, pathology, and even function [18]. Compared to other imaging modalities, MRI offers unique advantages including high resolution images, very good soft tissue contrast, and, unlike computed tomography (CT) or positron emission tomography (PET), it does not use ionizing radiation [19]. This makes it especially suitable for imaging fetuses, neonates and children, as magnetic fields are not harmful to living cells.

An alternative to MRI is cranial ultrasonography, a technique which is commonly used in neonatal care due to its low cost, portability, and lack of ionizing radiation [20]. However, ultrasound waves struggle to penetrate bone, which means that the ultrasound probe needs to be used through acoustic windows, such as the fontanelles (the gaps between the infant’s cranial bones), before they begin to close between 4–26 months of age [21]. Moreover, ultrasound is not sufficient to detect subtle changes in the brain’s anatomy throughout development, as its resolution and contrast between soft tissues is lower than that of MRI. For these reasons, MRI remains the gold standard for imaging the brain, due to its unique advantages, as well as its ability to image through the skull [22]. Furthermore, an MRI sequence can be sensitised to a wide range of morphological and physiological parameters, such as flow, diffusion, perfusion, blood oxygenation, and many others [23].

1.2.1 Structural MRI

Structural MRI is one of the most widely used imaging techniques for research and clinical purposes alike, as it provides good anatomical detail and a strong soft tissue contrast [19]. In neuroimaging studies, MRI can be used to non-invasively image the anatomy of the brain, making it possible to distinguish between different types of tissues, such as CSF, WM, and gray matter (GM).

In T_1 -weighted (T_1w) images of the adult brain, WM appears bright, *i.e.*, light gray (see Figure 1.3 young adult, magenta arrows), while CSF is void of signal, *i.e.*, black (Figure 1.3 young adult, cyan arrows), and GM has a medium intensity, *i.e.*, dark gray (Figure 1.3 young adult, green arrows). In contrast, T_2 -weighted (T_2w) images are inverted, with CSF appearing lightest in intensity, while WM is darkest (see Figure 1.3 bottom row).

Neonatal brains, however, differ significantly from the adult brain. For example, brain volumes differ vastly, with the neonatal brain ranging between 100 mL to 600 mL, while an average adult brain can be bigger than 1 L in volume [24, 25]. Moreover, differences exist in terms of the neonatal brain’s appearance in structural MRI when compared to the adult brain. For instance, even though cortical folding is already complete, the smaller resolution compared to anatomy results in more pronounced partial volume effects. Other changes, caused by brain maturation, during which the myelin sheath forms around WM tracts, cause an inversion of intensity distributions between WM and cortical gray matter (cGM) tissue types by the time the brain reaches adulthood. This is visible in Figure 1.3 where a 41 weeks PMA at scan neonate’s T_1w image exhibits high intensities in the cGM region (green arrows), and darker values in WM (magenta arrows). When the infant reaches 6 months of age, the cGM and WM intensities are close in value, while the adult brain shows WM having the highest intensity (magenta arrows), and cGM having a medium intensity (green arrows). This inversion of intensities is visible on the T_2w MR images as well, where the neonate’s cGM is darker than WM, while this reverses in the adult brain. CSF is dark at all ages in T_1w images, and bright in

T_2w MRI (cyan arrows).

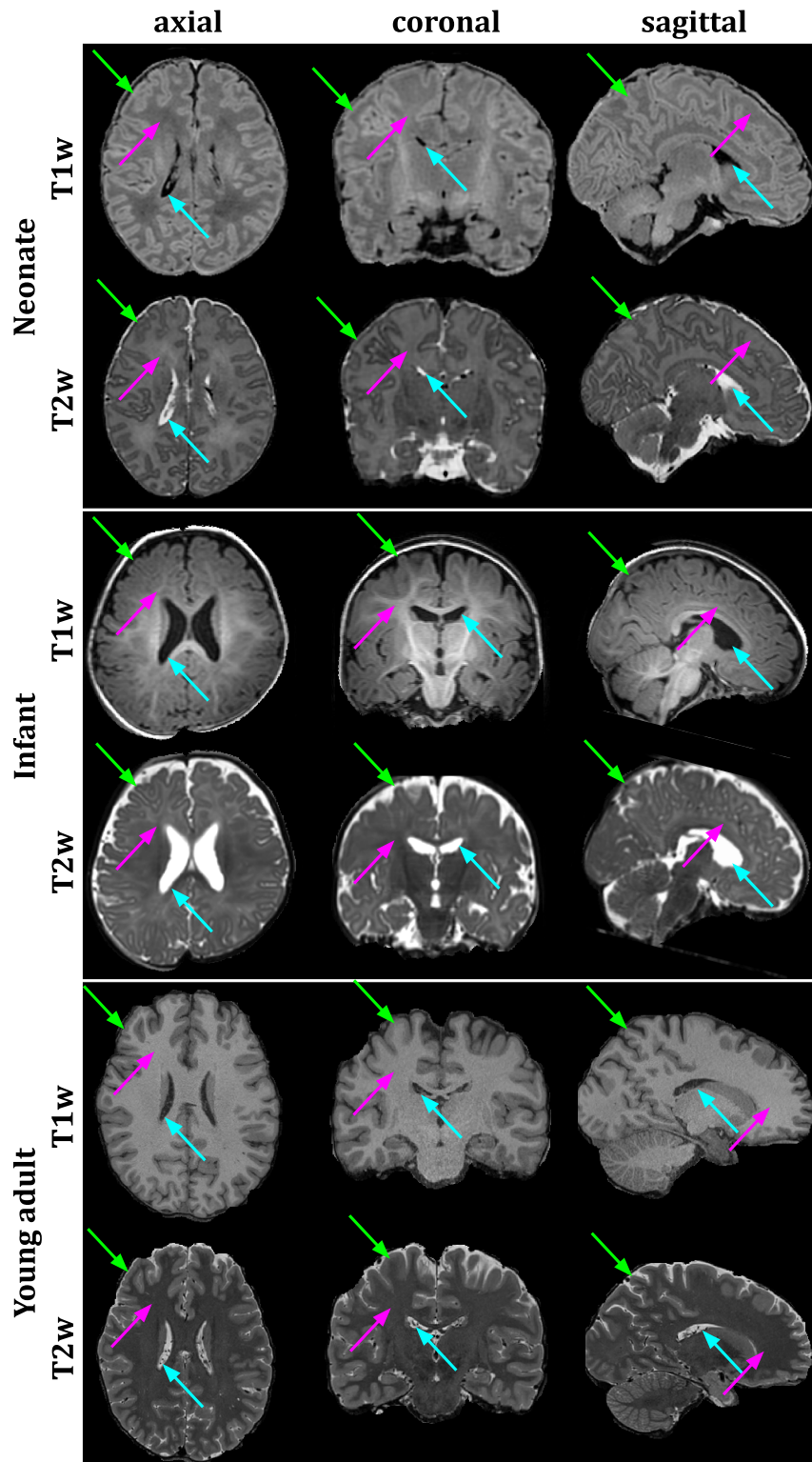


Figure 1.3: Axial, coronal and sagittal T_1w and T_2w MRI slices of a neonate (41 weeks PMA at scan), an infant (6 months), and a young adult (26-30 years). Green arrows point to regions of cortical gray matter (cGM), cyan arrows point to the cerebrospinal fluid (CSF) found inside the ventricles, while magenta arrows point to regions of white matter (WM).

1.2.2 Microstructural MRI

Microstructural imaging can be achieved with diffusion weighted magnetic resonance imaging (DW-MRI), an MRI modality which allows studying the diffusion characteristics of water molecules within tissue. For example, the properties of diffusion in brain WM fiber tracts can be described with the diffusion tensor (DT) model [26]. Moreover, several useful measures, such as fractional anisotropy (FA) [27] or orientation dependence of the diffusion, can be extracted from the DT model and studied. Figure 1.4 shows the mid-brain axial slices of five example neonates scanned at different ages (33.4 weeks, 37.1 weeks, 38.6 weeks, 41.1 and 43 weeks PMA at scan, respectively) and their corresponding T_2w , and diffusion tensor imaging (DTI) modalities.

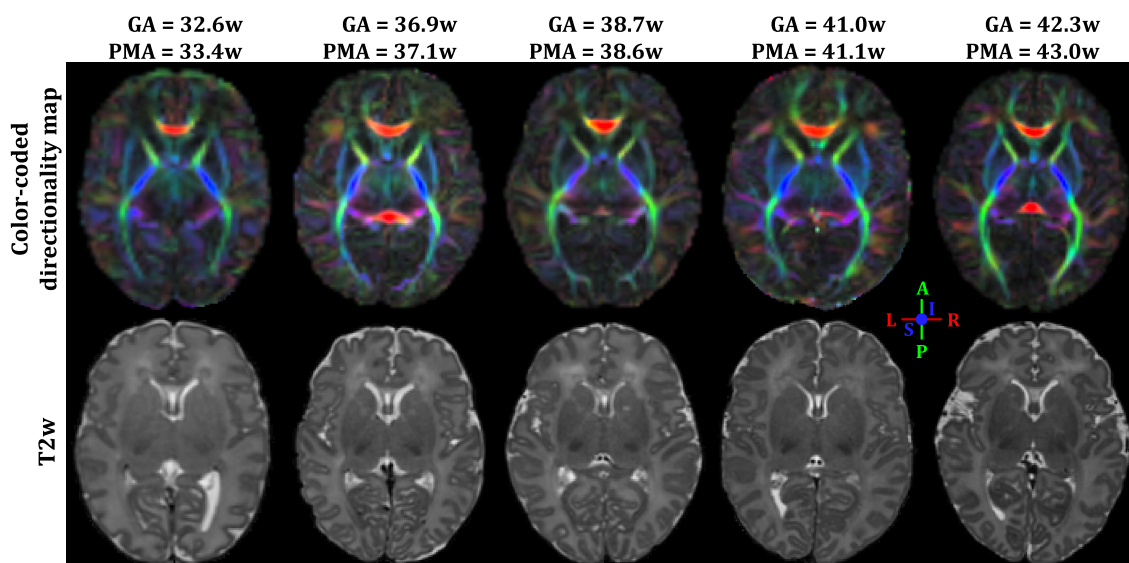


Figure 1.4: Axial slices of five neonates ranging from 33.4 weeks PMA at scan to 43 weeks PMA at scan, showing both the color-coded directionality map (DTI data weighted by their respective FA maps) and the T_2w images.

Diffusion Tensor. Free diffusion can be characterized by a single parameter D , known as the diffusion coefficient. However, biological tissue is rather complex, and water does not always diffuse freely as there are permeable and impermeable cellular structures which restrict it. When the surrounding barriers form coherent structures (such as WM axonal bundles), water will diffuse anisotropically, following patterns which reflect the neighboring structures. This is shown schematically in Figure 1.5 where free, unrestricted diffusion is shown as an isotropic diffusion ellipsoid, while restricted (but randomly organised) diffusion is shown as (also) an isotropic ellipsoid, but reduced in size. The final column shows an example of organised barriers where diffusion will be restricted perpendicular to the fiber bundle, and free parallel to it. In order to model water diffusion within biological structures, Basser *et al.* [26] introduced the diffusion tensor. This is a model of diffusion which can be completely characterized by a 3-by-3 symmetric positive-definite (SPD) matrix \mathbf{D}

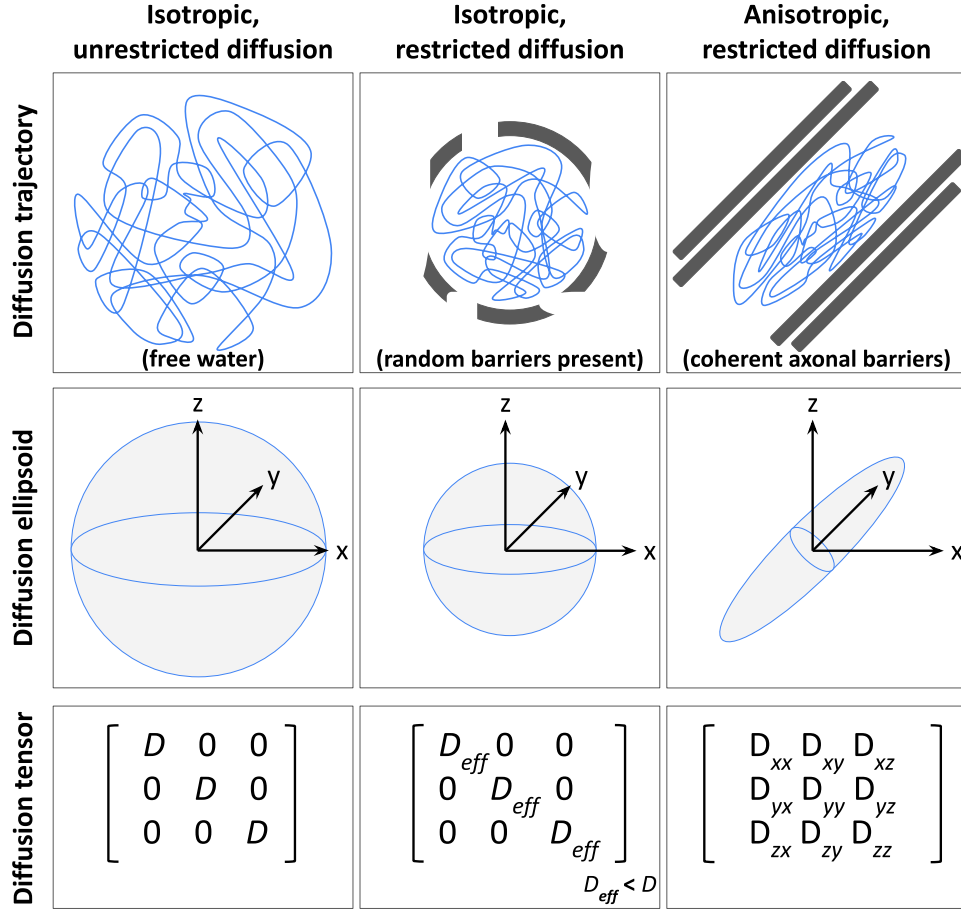


Figure 1.5: Example of diffusion trajectory, ellipsoids and tensors for isotropic unrestricted, isotropic restricted and anisotropic restricted diffusion. Image adapted from Mukherjee *et al.* [28].

with 6 independent components:

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix} \quad (1.1)$$

where $D_{yx} = D_{xy}$, $D_{zx} = D_{xz}$, and $D_{zy} = D_{yz}$. The last row of Figure 1.5 shows three examples of diffusion tensors for isotropic unrestricted ($D_{xx} = D_{yy} = D_{zz} = D$ and $D_{xz} = D_{xy} = D_{yz} = 0$), isotropic restricted ($D_{xx} = D_{yy} = D_{zz} = D_{eff} < D$) and anisotropic restricted diffusion.

The eigendecomposition of matrix \mathbf{D} always exists, and because it is positive-definite, its eigenvalues are positive. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the eigenvalues of \mathbf{D} and $\{\mathbf{e}_i\}_{i=1,2,3}$ its corresponding eigenvectors, then:

$$\mathbf{D} = \underbrace{\begin{bmatrix} e_{1x} & e_{2x} & e_{3x} \\ e_{1y} & e_{2y} & e_{3y} \\ e_{1z} & e_{2z} & e_{3z} \end{bmatrix}}_E \underbrace{\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}}_\Lambda \underbrace{\begin{bmatrix} e_{1x} & e_{2x} & e_{3x} \\ e_{1y} & e_{2y} & e_{3y} \\ e_{1z} & e_{2z} & e_{3z} \end{bmatrix}^{-1}}_{E^{-1}} \quad (1.2)$$

where $\mathbf{e}_i = (e_{ix}, e_{iy}, e_{iz})$. The eigenvector corresponding to the largest eigenvalue

(*i.e.*, \mathbf{e}_1) is assumed to be co-linear with the dominant fiber orientation within the voxel [29]. Figure 1.6 shows a schematic representation of an example diffusion tensor ellipsoid, together with the eigenvectors which define its orientation (*i.e.*, the principle axes of the ellipsoid), as well as the corresponding eigenvalues, which define its shape.

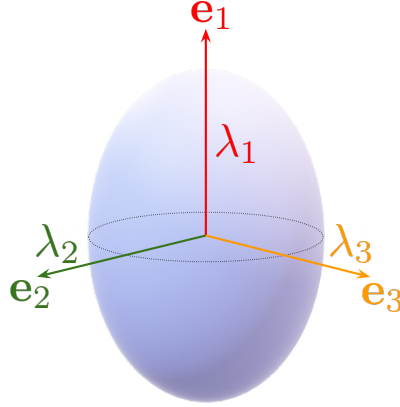


Figure 1.6: Schematic representation of the diffusion tensor ellipsoid. The principle axes are the three unit eigenvectors (\mathbf{e}_1 , \mathbf{e}_2 and \mathbf{e}_3), scaled by their corresponding eigenvalues (λ_1 , λ_2 and λ_3). Image adapted from Johansen *et al.* [29].

The relationship between the diffusion tensor ellipsoid, its eigenvalues, eigenvectors and its trace, is also exemplified in Figure 1.7, for different orientations of the same DT. The DT matrices are defined under each example ellipsoid, together with the corresponding eigenvalues (λ_1 , λ_2 and λ_3), eigenvectors (\mathbf{e}_1 , \mathbf{e}_2 and \mathbf{e}_3) and trace. Notice that, because the DT represents the orientation of diffusion in the laboratory frame of reference, the values in \mathbf{D} change as the ellipsoid is rotated (see Figure 1.7). At the same time, the three eigenvalues, as well as its trace, remain the same, as the shape of the tensor has not changed, while the eigenvectors change to reflect the new orientation of the tensor [30, 31].

DTI Scalars. By combining and weighting these eigenvalues, which are orientation invariant [32], one can derive different scalar-valued measures, and highlight specific features of water diffusion. Most commonly used DTI measures [33] include mean diffusivity (MD):

$$\text{MD} = \frac{1}{3}\text{Tr}(\mathbf{D}) = \frac{1}{3} \sum_{i=1}^3 \lambda_i \quad (1.3)$$

and FA [27, 34]:

$$\text{FA} = \sqrt{\frac{3}{2} \frac{\sum_{i=1}^3 (\lambda_i - \frac{1}{3}\text{Tr}(\mathbf{D}))^2}{\sum_{i=1}^3 \lambda_i^2}} \quad (1.4)$$

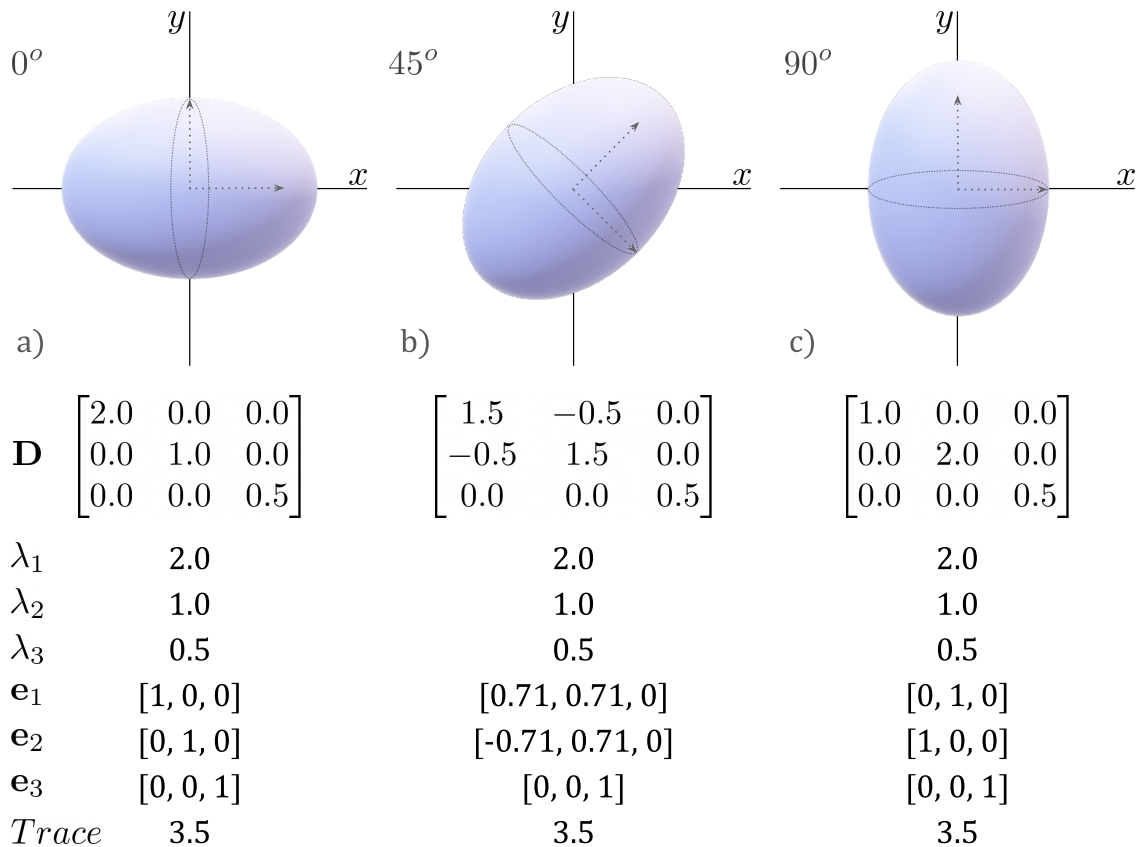


Figure 1.7: Schematic representation of the relationship between the diffusion tensor ellipsoid, its eigenvalues, eigenvectors and its trace. This example DT has a shape which can be described by its eigenvalues $\lambda_1 = 2.0$, $\lambda_2 = 1.0$ and $\lambda_3 = 0.5$, and it is being rotated about the z-axis (perpendicular to the plane) with 45° and with 90° , respectively. In a) the DT has its longest axis aligned with the x-axis and its middle axis aligned with the y-axis; in b) the ellipsoid has been rotated about the z-axis with 45° ; and in c) the ellipsoid is 90° rotated about the z-axis, such that its longest axis is aligned with the y-axis and its middle axis with the x-axis. Its eigenvalues (λ_1 , λ_2 and λ_3), eigenvectors (\mathbf{e}_1 , \mathbf{e}_2 and \mathbf{e}_3) and corresponding trace are shown underneath each example. Image adapted from Mori *et al.* [30].

1.2.3 Acquisition and preprocessing of neonatal MRI

MR acquisition of the neonatal brain poses unique challenges. For example, scanning times are often shorter than in adult MRI to limit the discomfort of the baby, but this results in reduced spatial and temporal resolution of the images. Even the clinical adult MR receiver head coil can pose problems in this cohort due to the smaller head sizes. Moreover, it is preferred that neonates are scanned without sedation, but MRI protocols, especially DW-MRI, are highly sensitive to head motion. At the same time, existing adult image analysis tools do not translate well to the neonatal cohort's image intensity distributions, which are highly variable within different weeks of development.

A state-of-the-art neonatal MRI dataset is the developing Human Connectome Project¹ (dHCP). It has brought many advances in both structural and diffusion MRI in order to achieve its aim of collecting high-quality imaging data of both preterm and term-born neonates. In fact, scans were acquired with advanced, optimised and bespoke methods for structural and microstructural images, and this dataset contains anatomical (T_1w , T_2w), resting state functional MRI (rsfMRI), and DW-MRI data, in both their original and after applying the processing pipelines described in Edwards *et al.* [11]. In addition, clinical, demographic and genetic information is also present, however these were not utilised in this thesis.

A second dataset, the Evaluation of Preterm Imaging² (ePrime) study [35], is an older dataset which did not benefit from advanced processing techniques. Similar to dHCP, the ePrime dataset also contains diffusion and functional MRI, clinical, demographic and genetic information, however these data were not utilised in this thesis. ePrime was focused on acquiring data from preterm-born neonates only, and its images did not go through motion correction or super-resolution reconstruction. This means that medical image analysis methods which were developed for the dHCP dataset cannot guarantee high quality predictions when applied to the ePrime data, as the source and target domains are dissimilar due to different acquisition protocols, or biases in patient cohorts. However, combining imaging data from multiple studies and sites is important to increase the sample size and thereby the statistical power of neuroimaging studies. Therefore, there is a need for harmonising MRI datasets in order to make sure that the differences arising from different image acquisition protocols do not affect the analysis performed on the combined data.

Figure 1.8 shows the mid-brain sagittal, coronal and axial slices of two example ePrime neonates scanned at 44.7 weeks PMA, and 43.6 weeks PMA, respectively, and two example dHCP neonates scanned at 44.2w PMA and 43.1 weeks PMA, respectively. The ePrime subjects showcase more motion corruption when compared to the dHCP neonates (see green arrows in Figure 1.8). Moreover, due to differences in acquisition, the ePrime data is generally more blurry, as can be seen by the variance of the Laplacian histogram plot shown in Figure 1.9. This measure is used to quantify the degree of blurriness in an image, as suggested by the survey paper of Pertuz *et al.* [36], with higher values corresponding to sharper images, and lower values corresponding to blurrier images [37].

The rest of this subsection is focused on describing the two neonatal datasets, dHCP [11] and ePrime [35], in terms of their respective acquisition protocols, as both these data are used throughout the thesis.

dHCP. Image acquisition of the dHCP dataset was undertaken at St. Thomas Hospital, London, on a Phillips 3 Tesla Achieva system (Philips Medical Systems, Best, The Netherlands), during natural sleep without sedation, and using a dedicated neonatal 32-channel receive coil with a custom-made acoustic hood [38]. To ac-

¹developingconnectome.org

²npeu.ox.ac.uk/prumhc/eprime-mr-imaging-177

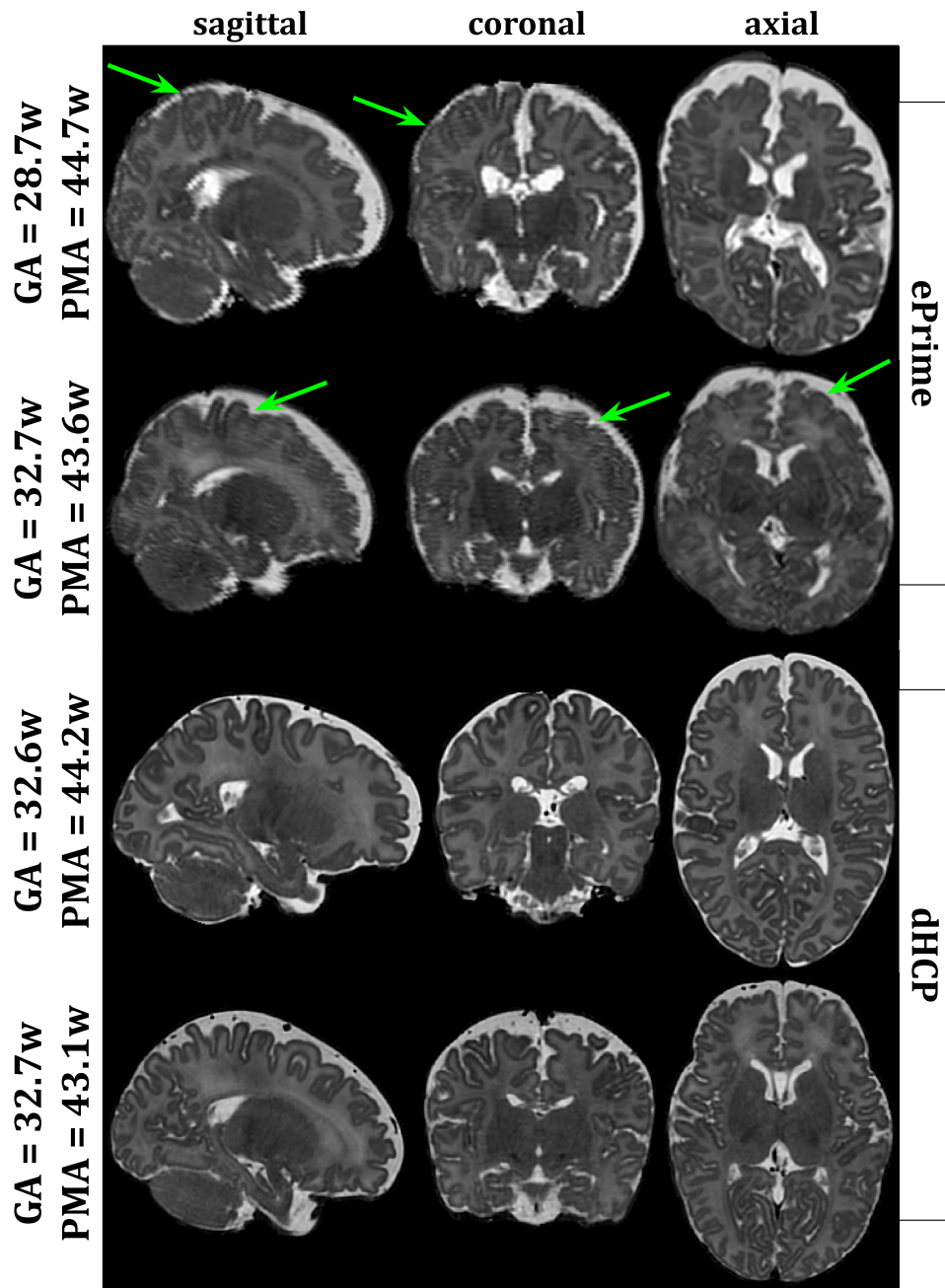


Figure 1.8: Example of two ePrime neonates scanned at 44.7 weeks PMA, and 43.6 weeks PMA, respectively, and two dHCP neonates scanned at 44.2w PMA and 43.1 weeks PMA, respectively. Green arrows point to regions where the ePrime data has visible motion artifacts.

quire T_2w structural images while reducing the effects of motion, the dHCP pipeline uses a turbo spin echo (TSE) sequence with parameters: repetition time $T_R = 12$ s, echo time $T_E = 156$ ms, overlapping slices with $0.8 \times 0.8 \times 1.6$ mm³ resolution, and SENSE factors of 2.11 for the axial plane and 2.58 for the sagittal plane. All data was motion corrected [39, 40] and resampled to an isotropic voxel size of 0.5 mm³.

For DW-MRI, the optimised imaging protocol uses a scattered slice multi-shell high angular resolution diffusion imaging (HARDI) acquisition strategy, coupled

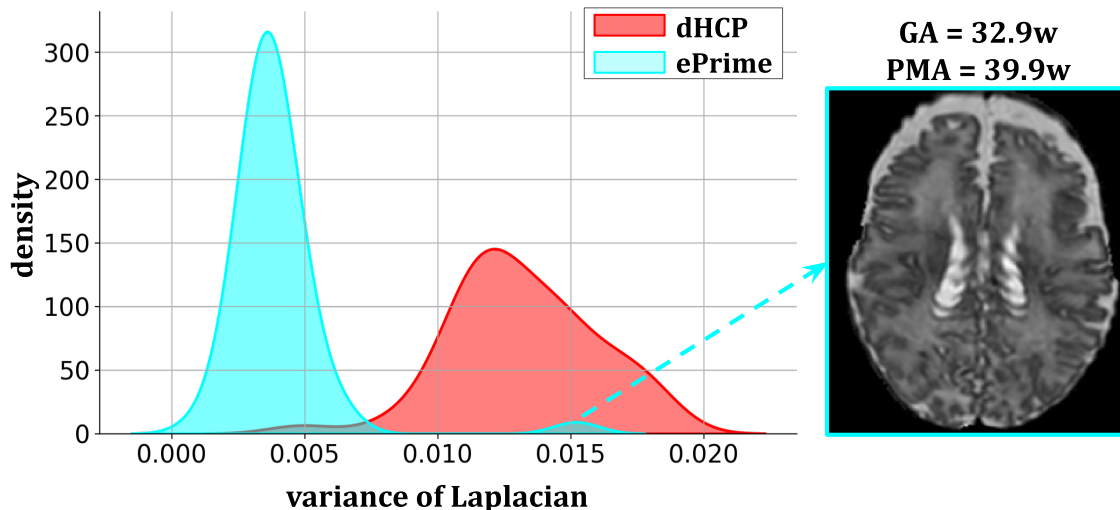


Figure 1.9: Density plot of variance of the Laplacian computed on the normalized intensities of T2w MRI volumes of both ePrime and dHCP neonatal datasets. Higher values correspond to sharper images, while lower values correspond to blurrier images. The ePrime peak shown with the cyan arrow corresponds to ePrime neonates with high degree of motion artifacts, such as the example infant whose axial image is displayed on the right hand side of the figure.

with a monopolar spin echo echo-planar imaging (SE-EPI) Stejskal-Tanner sequence ($\Delta = 42.5$ ms and $\delta = 14$ ms), to acquire 300 volumes in a short period of time (< 20 min) [41]. At the same time, the acquisition parameters of $T_R = 3.8$ s, $T_E = 90$ ms, a multiband factor of 4, a SENSE factor of 1.2, and a partial Fourier factor of 0.855, coupled with setting the slice thickness to 3 mm, the in-plane resolution to 1.5 mm, and the slice spacing to 1.5 mm, achieves optimal brain coverage in the presence of motion, while keeping a high signal-to-noise ratio (SNR) [41]. To acquire the 300 volumes, the diffusion gradients are sampled across 4 shells ($b = 0, 400, 1000,$ and 2600 s/mm² respectively), each with a different number of samples (20, 64, 88, and 128 samples respectively) [42]. Using thicker overlapping slices is essential to preventing gaps in data caused by motion, but can lead to blurred images in the through-plane direction. To solve this, a super-resolution reconstruction algorithm is applied [40] to the data to recover a 1.5 mm isotropic voxel resolution. Data preprocessing included image denoising [43], Gibbs-ringing artifact removal [44], and correction of magnetic field distortions through FSL Topup [45] which estimates the susceptibility and eddy currents-induced off-resonance maps. Finally, a slice-to-volume reconstruction framework which uses a bespoke spherical harmonics and radial decomposition (SHARD) method is applied to the data to correct subject motion and EPI distortions [46].

ePrime. The images which are part of the ePrime dataset were acquired with a Philips Intera 3T system and an 8-channel phased array head coil, using a T_2w TSE sequence with parameters: $T_R = 8.67$ s, $T_E = 160$ ms, and TSE factor 16. The in-plane resolution was set to 0.86×0.86 mm, and the slice thickness to 2 mm with an overlap of 1 mm. For each volume, the acquisition ranged between 92 and 106 slices in the transverse plane. Diffusion data was also acquired, however, this was

not utilised in the thesis.

1.3 Challenges in analysis of the neonatal brain MRI

The changes that the brain undergoes during the developmental period make models trained and tested on adult brain MRI not suitable for neonatal analysis. This, coupled with MR image artifacts induced by the small head sizes of this cohort which results in lower resolution images, as well as variable head sizes which can modulate the overall SNR, and also the inevitable head motion, make automatic segmentation or image registration of the neonatal brain a non-trivial problem.

Research into the developing brain has found links between MRI metrics and prematurity, clinical factors, and neurodevelopmental outcomes [47, 48, 49]. Studies have also shown that by combining structural and diffusion MR images one has the potential to better understand how the brain matures [50, 47]. Such analyses rely on accurate inter-subject or subject-to-template image registration methods. However, there is a lack of tools for combined analysis of structural and diffusion MRI in the same reference space [51], thus omitting to take into account the complementary information provided by using both.

The advancements brought to the acquisition and reconstruction protocols [39] have produced high-resolution T_1w and T_2w MR images. These modalities offer high contrast between different brain tissues and can delineate the cGM region well, but can suffer from varying intensities throughout development due to maturation processes [51]. For example, transient WM compartments (*e.g.*, periventricular regions) are highly heterogeneous during early brain development, as they change and evolve rapidly [52]. This results in intensity changes in the structural MR images (see Figure 1.10 green circles) consisting of a gradual darkening of the periventricular crossroads in the T_2w images, and a lightening in the T_1w images. These T_2w hyperintensities are reportedly due to a higher water content which gradually decreases with maturation causing the darkening [52].

Diffusion MRI is well suited to provide knowledge about the extent or location of well-aligned cytoarchitectures, with FA values remaining stable in the major WM tracts throughout neonatal development [53] (see Figure 1.10 magenta arrows). On the other hand, DW-MRI is sensitive to the decrease of FA values in the cortex [53] caused by the reduction of the radial orientation of the cortical architecture during development. In addition, FA maps provide much poorer delineation and lower contrast of cortex than structural MRI [51] (see Figure 1.10 cyan arrows).

It is therefore important to build tools that combine these complementary modalities to drive the registration process, which, in turn, will help with downstream

analysis. In fact, previous studies have shown that combining diffusion and structural data to drive the registration [54, 55, 56, 51, 57] improves the overall alignment. However, these approaches either weigh each channel similarly [54], or use pre-calculated certainty maps to highlight important regions [57, 56, 51].

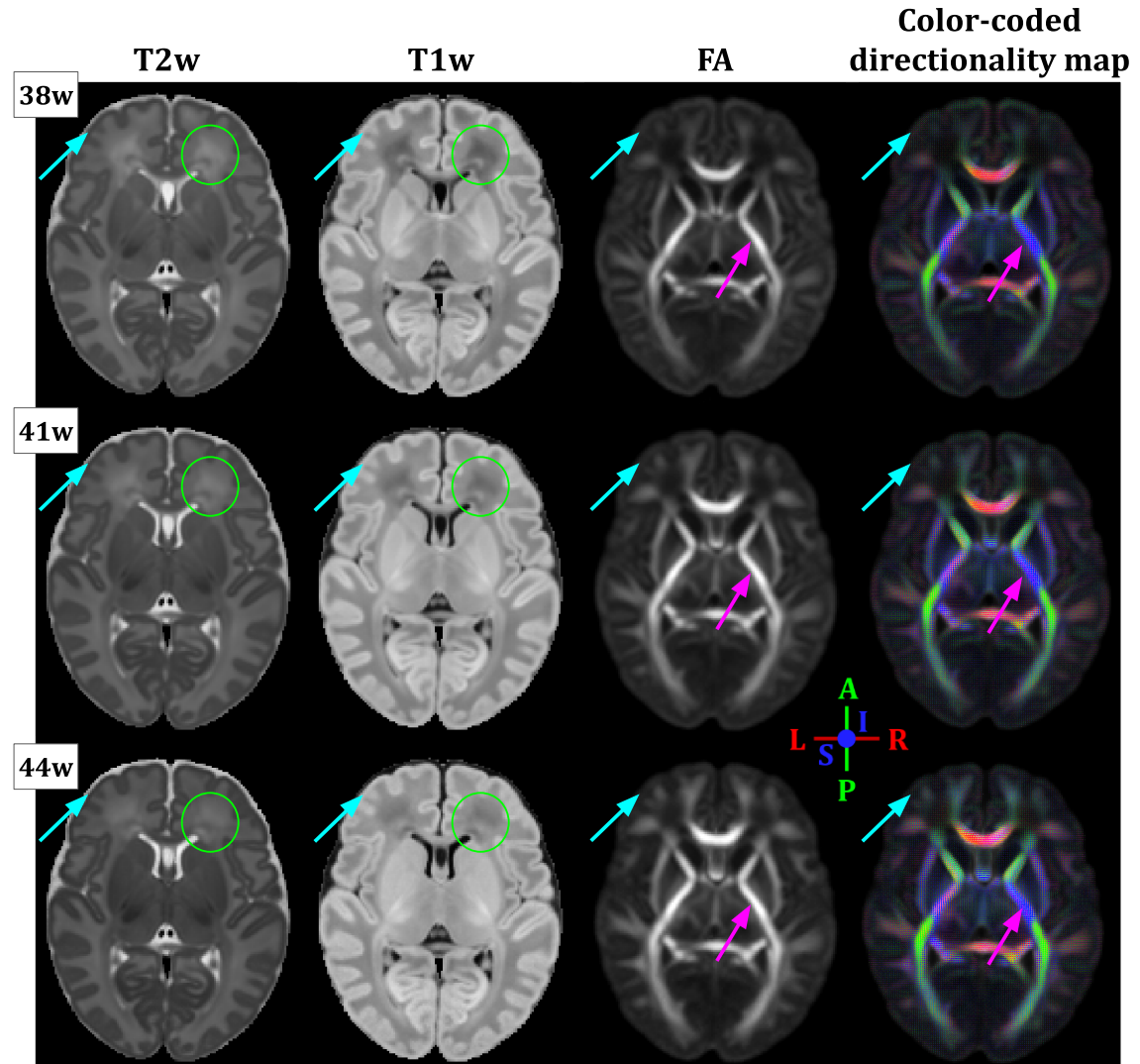


Figure 1.10: Axial slices of three neonatal atlases at 38 weeks, 41 weeks, and 44 weeks, respectively, showing both the structural data (T_2w and T_1w in the first two columns) and the microstructural data (FA maps and DTI data weighted by their respective FA maps in the last two columns). Green circles point to the periventricular crossroads, cyan arrows point to the cortical gray matter ribbon, and magenta arrows point to a region of a white matter tract known as the internal capsule. Image adapted from Uus *et al.* [51].

1.4 Thesis contributions

The main aim of this PhD project is to design deep learning algorithms suited for analysis of neonatal brain MRI. I aim to address the following challenges in neonatal brain MRI analysis that are not well solved by available neonatal segmentation and registration tools:

- *Deep learning models rely on the assumption that the source and target domains are drawn from the same distribution.* The performance of deep learning methods drops drastically when applied to images acquired with acquisition protocols or patient cohorts different than the ones used to train the models. At the same time, combining imaging data from multiple studies and sites is necessary to increase the sample size and thereby the statistical power of neuroimaging studies. Moreover, the lack of standardization in image acquisition protocols, scanner hardware, and software, can lead to inter-scanner variability, which has been demonstrated to affect measurements obtained for downstream analysis such as voxel-based morphometry [58], and lesion volumes [59]. Therefore, it is important to harmonize MRI datasets in order to ensure that the differences arising from various image acquisition protocols do not affect the analysis conducted on the combined data, where measures such as volumetric and cortical thickness should reflect brain anatomy and remain unaffected by the acquisition protocol or scanners utilized.
- *Leveraging multi-modal MRI for accurate alignment of the neonatal brain.* As a prerequisite for downstream tasks, accurate alignment of neonatal MRI of various modalities is needed. Structural and microstructural MRI modalities offer complementary information about morphology and tissue properties of the developing brain, which can be leveraged to achieve more accurate inter-subject alignment [54, 55, 56, 57]. This is particularly important during early life when the brain undergoes a rapid maturation process. Currently, the inter-subject alignment is most commonly driven by a single modality, such as structural [60] or diffusion [61]. Nevertheless, multi-channel image registration approaches that utilize multiple imaging modalities have been attempted, but they are based on simple averaging of the deformation fields from the individual channels [54], leading to solutions where both modalities are treated equally, or through weighting the deformation fields with pre-calculated spatial gradient maps [57, 56, 51], where the local weights are fixed for the entire image domain.

For these reasons, I propose:

- To investigate domain adaptation (DA) methods with the aim of predicting brain tissue segmentations of T_2w MRI data of an unseen neonatal population.

- To develop an attention-based deep learning registration model that learns a spatially varying importance map for each individual channel (modality), thus taking advantage of their complementary nature.
- To develop a deep-learning based multi-channel registration network that combines intensity-based registration of structural data with metrics that align white matter tracts in diffusion data.

1.5 Thesis outline

My thesis is composed of 7 chapters. **Chapter 1** provides introduction to neonatal brain MRI image analysis. **Chapters 2** and **3** form the background and literature review of this thesis, with a focus on medical image registration and segmentation. The novel contributions of this thesis are described in **Chapters 4–5–6**. The thesis conclusions and future work are detailed in **Chapter 7**.

Chapter 1, the current chapter, introduces the reader to the neonatal brain and the particularities of the highly versatile MR imaging modality used to non-invasively study it. Moreover, this first chapter describes two neonatal databases, dHCP and ePrime, which are used throughout the thesis, introduces the main challenges in analysis of neonatal brain MRI, and proposes deep learning based solutions as the main contributions of this research.

Chapter 2 starts by presenting the relevant theoretical background for classical MR image registration of both structural and microstructural data, and discusses the main methods used in the literature. More specifically, it lays the groundwork for this thesis by introducing the relevant similarity measures or penalty terms used to train our proposed registration networks, as well as the theoretical background for understanding the particularities of registering higher order diffusion data. In the second half of this chapter, registration-based and intensity-based medical image segmentation approaches are presented, with a particular focus on segmentation of brain MRI. Moreover, it presents the most common segmentation frameworks, including the algorithm used to produce the neonatal brain tissue and structure labels used for training and evaluation throughout this thesis. The chapter concludes with a discussion of the most common evaluation metrics for validating image segmentation models, with a focus on spatial overlap and surface based metrics.

Chapter 3 focuses on deep learning techniques, starting from the main theoretical building blocks of artificial neural networks. A survey of state-of-the-art deep learning based medical image registration and segmentation techniques follows, with the aim of both complementing the classic literature presented in the previous chapter, as well as discussing some of the baseline models used throughout this thesis. The chapter ends with a discussion of more advanced deep learning techniques, such as visual attention, where the literature is surveyed in terms of channel, spatial,

mixed, and non-local attention, and domain adaptation methods, grouped into supervised, semi-supervised and unsupervised techniques. Methods related to deep learning visual attention, as well as unsupervised domain adaptation, are used in the results chapters.

Chapter 4 presents a study where unsupervised DA methods are investigated, with the aim of harmonizing brain tissue segmentation maps of dHCP and ePrime cohorts. Here, we find that adversarial domain adaptation in the image space performs best for our target dataset. Moreover, as a proof-of-principle, we show the importance of harmonising the cortical tissue maps by investigating the association between neonatal cortical thickness and a language outcome measure.

Chapter 5 demonstrates a novel attention-based multi-channel deep learning image registration framework which improves the alignment of datasets consisting of neonatal T_2w MRI and diffusion weighted imaging (DWI)-derived FA maps. We compare the proposed method with models trained on single- or multi-channel data, as well as introducing channel and spatial attention blocks throughout the registration network. Our main results show that combining the two complementary modalities is best achieved with the use of a global weight which balances the two channels, as well as the locally varying spatial attention map.

Chapter 6 extends the multi-channel deep learning registration network to work with higher-order DTI data. For this, layers which account for the change in orientation of diffusion tensors induced by the predicted deformation field are added to the network. Moreover, an evaluation of the accuracy of the white matter alignment shows improvements which cannot be achieved without the use of the higher order DTI data.

Chapter 7 concludes with a summary of the thesis contributions, as well as highlights limitations and discusses future avenues.

Medical image registration and segmentation for neonatal brain MRI

This chapter offers an overview of medical image analysis techniques, starting with the theory and applications of medical image registration in **Section 2.1**, and followed by approaches for detecting and segmenting tissue structures in medical images in **Section 2.2**.

Section 2.1 focuses on presenting the main concepts important for understanding medical image registration methods, with a focus on image similarity measures (**Section 2.1.1**), regularisation penalties (**Section 2.1.2**), transformation models (**Section 2.1.3**), and the particularities of registering higher order DT-MR data (**Section 2.1.4**).

Section 2.2 focuses on presenting the main concepts important for understanding medical image segmentation methods, with a focus on describing registration- and intensity-based approaches (**Sections 2.2.1** and **2.2.2**), followed by an overview of medical image segmentation frameworks (**Section 2.2.3**), and ends with a description of evaluation metrics for medical image segmentation applications (**Section 2.2.4**), respectively.

2.1 Medical image registration

Image registration is the field of medical image analysis which focuses on establishing anatomical correspondences between two or more images of tissues or organs. The key terms used in medical image registration are:

- *pairwise* - registration of two images ; *groupwise* - registration of more than two images
- *mono-modal* - registration of images acquired using the same image modality and acquisition parameters; *multi-modal* - registration of images acquired through different modalities
- *intra-subject* - registration of images representing the anatomy of the same subject ; *inter-subject* - registration of images representing the anatomy of different subjects
- *multi-channel* - registration of multiple modality images of the same anatomy for each subject

The purpose of an image registration algorithm is to find an alignment between two or more images such that corresponding features can be related. Mathematically, image registration is typically formulated as an optimization problem where the applied transformation is iteratively optimized based on an energy function.

Let $\mathbf{F} : \mathbb{R}^d \rightarrow \mathbb{R}$, $\mathbf{M} : \mathbb{R}^d \rightarrow \mathbb{R}$ represent two d -dimensional scalar-valued volumes known as the *fixed (target)* and the *moving (source)* images, respectively, and let $\varphi(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$ denote the deformation field φ at the spatial coordinate (position vector) $\mathbf{x} \in \mathbb{R}^d$, where \mathbf{u} is the continuous displacement field [62]. Then, the optimization problem becomes:

$$\hat{\varphi} = \arg \min_{\varphi} \mathcal{L}(\mathbf{F}, \mathbf{M}(\varphi)) \quad (2.1)$$

where

$$\mathcal{L}(\mathbf{F}, \mathbf{M}(\varphi)) = \mathcal{L}_{dsim}(\mathbf{F}, \mathbf{M}(\varphi)) + \lambda \mathcal{L}_{smooth}(\varphi) \quad (2.2)$$

is the energy (loss) function. $\mathcal{L}(\mathbf{F}, \mathbf{M}(\varphi))$ is composed of a dissimilarity measure \mathcal{L}_{dsim} between the warped moving image $\mathbf{M}(\varphi)$ and the fixed image \mathbf{F} and a smoothness constraint \mathcal{L}_{smooth} imposed on the deformation field φ . Here, λ is the regularization parameter.

The deformation field $\varphi : \Omega_{\mathbf{F}} \rightarrow \Omega_{\mathbf{M}}$ is a mapping between the spatial coordinates which are part of the fixed image domain ($\Omega_{\mathbf{F}}$) to the respective spatial coordinates of the moving image domain ($\Omega_{\mathbf{M}}$). This is visually represented in Figure 2.1, where $\mathbf{x} \in \Omega_{\mathbf{F}}$, and $\varphi(\mathbf{x}) \in \Omega_{\mathbf{M}}$. This can be achieved for both intensity based images, as well as surface-based, point-based or tensor-based applications. For images based

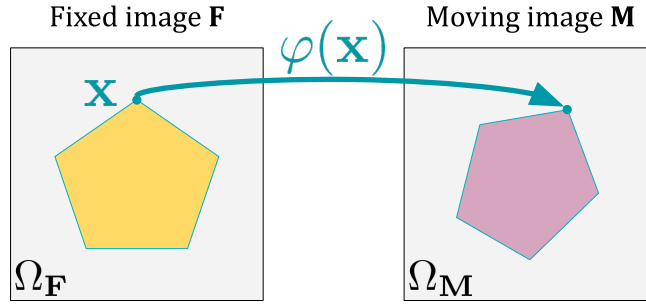


Figure 2.1: Example of a spatial mapping φ between fixed and moving images. The location $\mathbf{x} \in \Omega_{\mathbf{F}}$ in the fixed image \mathbf{F} is mapped through the deformation φ to its corresponding location in the moving image \mathbf{M} , at position $\varphi(\mathbf{x}) \in \Omega_{\mathbf{M}}$.

on the latter type, an extra step is required to reorient the tensors in accordance with the local deformation (see Section 2.1.4 for more details).

The dissimilarity measure \mathcal{L}_{sim} can be based on differences in intensity or tensor values, on correlations, or on the amount of shared information between the two images. Cross-correlation and mutual information metrics are often used when the two volumes have varying intensity distributions; for example, when the two images were acquired with different medical imaging modalities. Lastly, the smoothness constraint \mathcal{L}_{smooth} can be used to regularize the transformation such that it favours specific properties of the solution.

Thus, a generic registration algorithm is composed of the following steps:

1. **Transformation model.** First, the source image \mathbf{M} is transformed to the target image domain ($\Omega_{\mathbf{F}}$) using a **transformation model**, which can be either global or non-rigid. Some example global transformations can be seen in Figure 2.2.
2. **Interpolation/Resampling.** After the transformation φ is applied to the source image, the intensity values of the transformed image do not necessarily map to the discrete locations of the target image domain (see Figure 2.3). For this reason, different **interpolation strategies** can be used to retrieve the values at the specific locations. Some popular strategies, such as nearest neighbours, linear or cubic spline interpolation, are shown in Figure 2.4 for the 1-dimensional case.
3. **Similarity measure.** The **similarity** between the transformed moving image and the fixed image is then assessed. Different measures of similarity have been used in the literature and, for scalar data, can be classified in two categories: feature-based and intensity-based. Feature-based similarity measures require a pre-processing step in order to extract useful features (points, lines, surfaces) which are then used to compute the similarity. Intensity-based measures are computed from the voxel intensities directly. When aligning non-scalar data, such as DTI or HARDI extracted microstructural orientation distribution functions (ODF) data, the similarity measures are derived

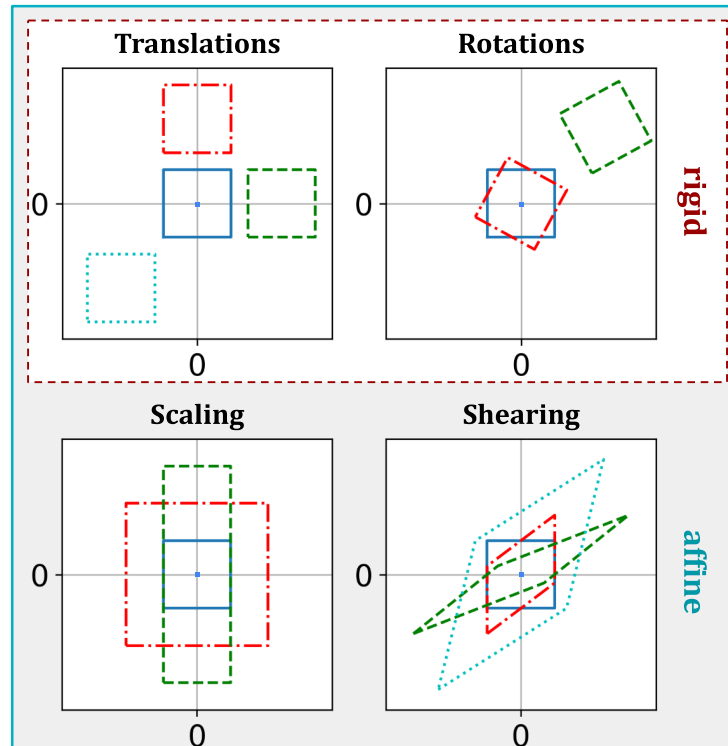


Figure 2.2: 2D global transformations of a blue square initially centered around $(0, 0)$. First row shows **rigid transformations**, such as translations: in y (red dashed), in x (green dashed), and in both x and y (cyan dotted), as well as rotations: about z (red dashed), or rotated and translated (green dashed). The second row shows **affine transformations**, such as scaling: along the y axis (green dashed), or in both directions (red dashed), and shearing: along the y axis (red dashed), along both x and y axis (green dashed), or sheared in both directions and scaled in both x and y (cyan dotted). Note that **affine transformations** include both rotations and translations, on top of scaling and shearing.

from the higher-order data. Moreover, when dealing with multi-modal data, statistics-based or information-based techniques can be used.

4. **Regularisation.** As registration is an ill-posed problem, it is often the case that constraints are added to the transformation model to produce realistic deformations. This practice is referred to as **regularisation** and can be *explicit* or *implicit* depending on where the regularisation is applied [63]. **Explicit** regularisation is applied to the deformation model, where some algorithms take advantage of physics-based properties [64, 65], others use composition schemes to generate topology preserving deformation fields [66], or use simpler methods such as smoothing [67]. **Implicit** regularisation takes the form of a penalty term which is added to the loss function. This penalty can be used to promote smooth and realistic deformation models [68, 69].
5. **Optimiser.** Finally, an **optimiser** is used in order to minimize the value of the dissimilarity measure by changing the transformation model's parameters. The following subsections will provide an overview of the most commonly used dissimilarity metrics and transformation models.

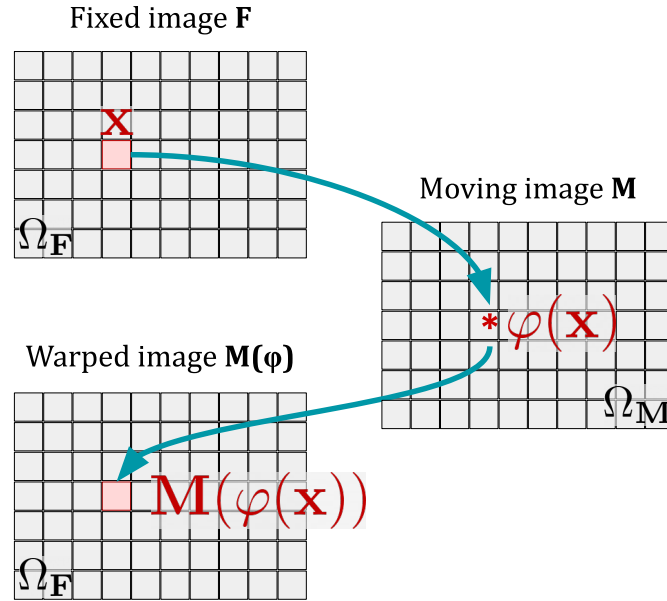


Figure 2.3: Illustration of image resampling in the context of image registration, showing how location $\mathbf{x} \in \Omega_{\mathbf{F}}$ in the fixed image \mathbf{F} is mapped through the deformation φ to its corresponding location in the moving image \mathbf{M} , at position $\varphi(\mathbf{x}) \in \Omega_{\mathbf{M}}$. During the resampling step, the warped image $\mathbf{M}(\varphi(\mathbf{x}))$ is created by interpolating the intensities of the moving image \mathbf{M} at the locations determined by the space of the fixed image \mathbf{F} .

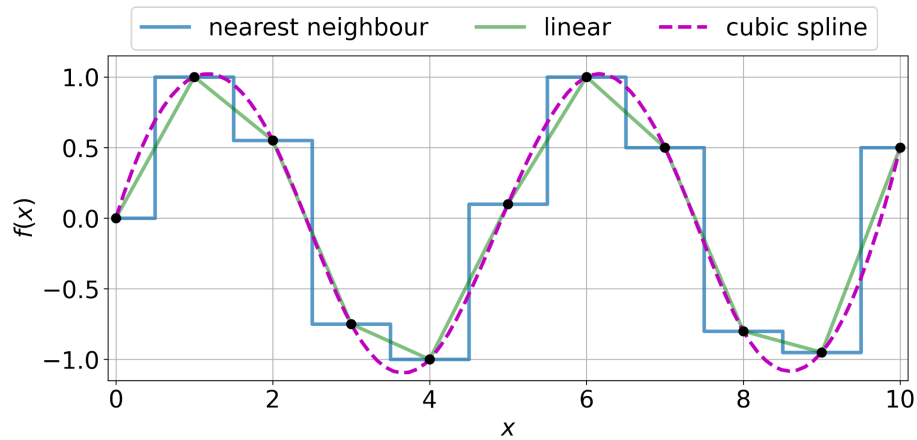


Figure 2.4: Example of different 1D interpolation strategies used to connect the sample points (shown in black). Nearest neighbours, linear and cubic spline interpolations are amongst the most popular types.

2.1.1 Image similarity measures

In this subsection, we summarize the most commonly used image similarity measures for medical image registration applications. All of the metrics presented below are written as a *dissimilarity* measure, meaning that an optimisation algorithm will have to minimise the metrics to obtain a good alignment (\mathcal{L}_{dissim} in equation 2.2). Moreover, this section focuses on scalar data only, such as T_1w , T_2w , or diffusion-

derived scalar-valued measures (*e.g.*, FA or MD maps). For higher-order data, see Section 2.1.4.

Intensity differences

First, when dealing with **mono-modal** data, some dissimilarity measures directly compare image intensities. For example, the simplest measure that can be used in a mono-modal image registration application is the **sum of squared differences (SSD)**. Also known as the mean squared error (MSE), the SSD is defined as:

$$\mathcal{D}_{\text{SSD}}(\mathbf{F}, \mathbf{M}(\varphi)) = \frac{1}{N} \sum_{\mathbf{x} \in \Omega_{\mathbf{F}}} |\mathbf{F}(\mathbf{x}) - \mathbf{M}(\varphi(\mathbf{x}))|^2 \quad (2.3)$$

where N is the number of voxels in the $\Omega_{\mathbf{F}}$ domain. In fact, SSD is an optimal measure when the images differ by only Gaussian noise [70, 71]. In the literature, the SSD has been widely used, with a few notable examples by Hajnal *et al.* [72, 73], and Friston *et al.* [74, 75].

The SSD measure is, however, sensitive to outliers. To overcome this limitation, the **sum of absolute differences (SAD)** measure can be used instead. Mathematically, it is defined as:

$$\mathcal{D}_{\text{SAD}}(\mathbf{F}, \mathbf{M}(\varphi)) = \frac{1}{N} \sum_{\mathbf{x} \in \Omega_{\mathbf{F}}} |\mathbf{F}(\mathbf{x}) - \mathbf{M}(\varphi(\mathbf{x}))| \quad (2.4)$$

As the SSD and SAD measures rely on the assumption of similar image intensity distributions for the same structures, they are not suitable for **multi-modal** image registration applications.

Correlation techniques

Correlation techniques relax this by assuming a linear relationship between the intensity values in the images [63]. The **cross correlation (CC)**, initially used by Lewis *et al.* [76] for 2D image matching, is one such measure, and it is defined as:

$$\mathcal{D}_{\text{CC}}(\mathbf{F}, \mathbf{M}(\varphi)) = -\frac{1}{N} \sum_{\mathbf{x} \in \Omega_{\mathbf{F}}} \mathbf{F}(\mathbf{x}) \cdot \mathbf{M}(\varphi(\mathbf{x})) \quad (2.5)$$

A more broadly used measure, however, is the **normalised cross correlation (NCC)** (also known as the *correlation coefficient* [77]), which first subtracts the average intensity from the images and divides by the standard deviation. It is defined as:

$$\mathcal{D}_{\text{NCC}}(\mathbf{F}, \mathbf{M}(\varphi)) = -\frac{1}{N} \frac{\sum_{\mathbf{x} \in \Omega_{\mathbf{F}}} (\mathbf{F}(\mathbf{x}) - \bar{F}) \cdot (\mathbf{M}(\varphi(\mathbf{x})) - \bar{M})}{\sqrt{\sum_{\mathbf{x} \in \Omega_{\mathbf{F}}} (\mathbf{F}(\mathbf{x}) - \bar{F})^2 \cdot \sum_{\mathbf{x} \in \Omega_{\mathbf{F}}} (\mathbf{M}(\varphi(\mathbf{x})) - \bar{M})^2}} \quad (2.6)$$

where \overline{F} is the mean voxel value in the fixed image \mathbf{F} within the Ω_F domain, \overline{M} is the mean voxel value in the transformed moving image $\mathbf{M}(\varphi(\mathbf{x}))$ within the same domain.

When locally varying intensities exist, a more robust measure is the **local normalised cross correlation (LNCC)** [60, 78, 79], which computes local means over smaller image regions. Mathematically it is defined as:

$$\mathcal{D}_{\text{LNCC}}(\mathbf{F}, \mathbf{M}(\varphi), \omega_{\mathbf{F}}) = -\frac{1}{N_{\omega}} \frac{\sum_{\mathbf{x} \in \omega_{\mathbf{F}}} (\mathbf{F}(\mathbf{x}) - \overline{F}) \cdot (\mathbf{M}(\varphi(\mathbf{x})) - \overline{M})}{\sqrt{\sum_{\mathbf{x} \in \omega_{\mathbf{F}}} (\mathbf{F}(\mathbf{x}) - \overline{F})^2 \cdot \sum_{\mathbf{x} \in \omega_{\mathbf{F}}} (\mathbf{M}(\varphi(\mathbf{x})) - \overline{M})^2}} \quad (2.7)$$

where $\omega_{\mathbf{F}} \subset \Omega_{\mathbf{F}}$ is a small image region centred at $\mathbf{x} \in \Omega_{\mathbf{F}}$, and N_{ω} is the number of voxels contained in the sub-volume. Equation 2.7 is averaged over the whole image domain $\Omega_{\mathbf{F}}$.

Information theoretic techniques

For multi-modal applications, the intensities do not generally have a linear relationship. In this case, the more suitable image dissimilarity techniques pertain to the information theoretic class [70]. One such measure of information is the Shannon's formula for entropy [80]. The entropy for the fixed image \mathbf{F} is defined as:

$$\mathbf{H}(\mathbf{F}) = - \sum_{f \in \mathbf{F}(\mathbf{x})} p(f) \log(p(f)) \quad (2.8)$$

where $\mathbf{F}(\mathbf{x})$ is the set of image intensity values at each location $\mathbf{x} \in \Omega_{\mathbf{F}}$. The entropy is usually estimated from the histogram of the image, or through a Parzen window approach [81, 82].

For image registration applications, the **joint entropy (JE)** of the two images is minimized [83, 84]. Defined as:

$$\begin{aligned} \mathcal{D}_{\text{JE}}(\mathbf{F}, \mathbf{M}(\varphi)) &= \mathbf{H}(\mathbf{F}, \mathbf{M}(\varphi)) \\ &= - \sum_{f \in \mathbf{F}(\mathbf{x})} \sum_{m \in \mathbf{M}(\varphi(\mathbf{x}))} p(f, m) \log(p(f, m)) \end{aligned} \quad (2.9)$$

where $p(f, m)$ represents the probability of having intensity f in image \mathbf{F} and intensity m in image \mathbf{M} at the same spatial location, the aim of the JE is to reduce the dispersion of the joint probability distribution. Similarly to the entropy \mathbf{H} , JE can be computed from the joint histogram of the two images.

One drawback of the JE is that it is possible to optimize it by reducing the content of either image [70], or when only the background regions overlap [85]. To solve this, Viola *et al.* [71] and Maes *et al.* [86] introduced the **mutual information**

(**MI**), which is computed based on the JE and the marginal entropies:

$$\begin{aligned} \mathcal{D}_{\text{MI}}(\mathbf{F}, \mathbf{M}(\varphi)) &= -\text{MI}(\mathbf{F}, \mathbf{M}(\varphi)) \\ &= -\left(\mathbf{H}(\mathbf{F}) + \mathbf{H}(\mathbf{M}(\varphi)) - \mathbf{H}(\mathbf{F}, \mathbf{M}(\varphi))\right) \end{aligned} \quad (2.10)$$

where the marginal entropy for the fixed image $\mathbf{H}(\mathbf{F})$ is defined in equation 2.8, and the marginal entropy for the transformed moving image is:

$$\mathbf{H}(\mathbf{M}(\varphi)) = - \sum_{m \in \mathbf{M}(\varphi(\mathbf{x}))} p(m) \log(p(m))$$

As the marginal entropies \mathbf{H} have to be maximised, the MI penalizes solutions where only the background regions overlap [85].

Finally, the normalised version of MI was introduced by Studholme *et al.* [87]. Known as the **normalised mutual information (NMI)**, it is defined as:

$$\begin{aligned} \mathcal{D}_{\text{NMI}}(\mathbf{F}, \mathbf{M}(\varphi)) &= -\text{NMI}(\mathbf{F}, \mathbf{M}(\varphi)) \\ &= -\frac{\mathbf{H}(\mathbf{F}) + \mathbf{H}(\mathbf{M}(\varphi))}{\mathbf{H}(\mathbf{F}, \mathbf{M}(\varphi))} \end{aligned} \quad (2.11)$$

and was shown to be more robust to variations in image overlap [87].

2.1.2 Image registration regularisation through penalty terms

In this subsection, we summarize the most commonly used image registration regularisation penalties. These measures are added to the overall loss function to be minimized in order to constrain the predicted transformation model to realistic deformations (see $\mathcal{L}_{\text{smooth}}$ in equation 2.2).

One such penalty that can be added to the overall energy function is the **diffusion regulariser** [88, 89], defined as sum of the norm of the gradients of the transformation in each dimension:

$$\mathcal{L}_{\text{diff}}(\mathbf{u}) = \int_{\Omega} \|\nabla \mathbf{u}\|_2^2 d\Omega \quad (2.12)$$

where \mathbf{u} is the continuous displacement field of the spatial mapping $\varphi(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$.

Wahba *et al.* [90] (in 2D) and Rueckert *et al.* [68] (in 3D) introduce a regularisation strategy based on a **bending energy (BE)** penalty term, which is defined as:

$$\begin{aligned} \mathcal{L}_{\text{BE}}(\varphi) &= \frac{1}{V} \int_{\Omega} \left(\frac{\partial^2 \varphi}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \varphi}{\partial y^2} \right)^2 + \left(\frac{\partial^2 \varphi}{\partial z^2} \right)^2 + \\ &\quad 2 \left(\frac{\partial^2 \varphi}{\partial xy} \right)^2 + 2 \left(\frac{\partial^2 \varphi}{\partial xz} \right)^2 + 2 \left(\frac{\partial^2 \varphi}{\partial yz} \right)^2 d\Omega \end{aligned} \quad (2.13)$$

where V is the volume of the image domain. In this thesis, we use the **bending energy** penalty when training the proposed image registration neural networks.

Finally, a transformation model can also be constrained to be incompressible. This was introduced by Rohlfing *et al.* [91] through a penalty term based on the Jacobian of the transformation at every voxel. As the Jacobian determinant is related to the local change in volume (with a value of 1 representing no change, a value smaller than 1 representing shrinkage and a value larger than 1 representing expansion), penalising it will regularise the transformation. The Jacobian matrix of the deformation field φ at location \mathbf{x} is defined as:

$$\mathbf{JAC}(\varphi(\mathbf{x})) = \begin{bmatrix} \frac{\partial \varphi_x(\mathbf{x})}{\partial x} & \frac{\partial \varphi_x(\mathbf{x})}{\partial y} & \frac{\partial \varphi_x(\mathbf{x})}{\partial z} \\ \frac{\partial \varphi_y(\mathbf{x})}{\partial x} & \frac{\partial \varphi_y(\mathbf{x})}{\partial y} & \frac{\partial \varphi_y(\mathbf{x})}{\partial z} \\ \frac{\partial \varphi_z(\mathbf{x})}{\partial x} & \frac{\partial \varphi_z(\mathbf{x})}{\partial y} & \frac{\partial \varphi_z(\mathbf{x})}{\partial z} \end{bmatrix} \quad (2.14)$$

Sdika *et al.* [92] introduced into their optimisation scheme a constraint on a positive Jacobian determinant of the deformation field, while Rohlfing *et al.* [91] and Modat *et al.* [69] propose to penalise the log-transformed Jacobian determinant. Other regularisation penalties exist, such as the sum of the Laplacian of the deformation model [89], but are out of scope for this thesis.

2.1.3 Transformation models

In this subsection we summarize the transformation models used in medical image registration, starting from global transformations, and ending with non-rigid transformations. At the end of the section, the focus will turn to the most common intensity-based non-rigid registration methods, grouped by the choice of the transformation model (either parametric or non-parametric).

Global transformations

First, **global transformations** are mappings where all the voxels in the warped image are transformed using a single transformation model [70]. These can be either *rigid* or *affine*, with a few examples shown in Figure 2.2.

Rigid transformations are mappings that preserve length and are made up of translations, rotations, and a combination of the two. In 3D space, rigid deformations can be represented with a single square matrix as the product of 3 rotation matrices and 3 translation matrices:

$$M_{rig} = \underbrace{R(\theta_x, \theta_y, \theta_z)}_{R_x(\theta_x)R_y(\theta_y)R_z(\theta_z)} Tr(t_x, t_y, t_z)$$

where:

$$R_x(\theta_x) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x & 0 \\ 0 & \sin \theta_x & \cos \theta_x & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad R_y(\theta_y) = \begin{pmatrix} \cos \theta_y & 0 & \sin \theta_y & 0 \\ 0 & 1 & 0 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R_z(\theta_z) = \begin{pmatrix} \cos \theta_z & -\sin \theta_z & 0 & 0 \\ \sin \theta_z & \cos \theta_z & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad Tr(t_x, t_y, t_z) = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with $\theta_x, \theta_y, \theta_z$ and t_x, t_y, t_z being the angles of rotation and the translations along each axis, thus being parametrised by 6 degrees of freedom.

Affine transformations are global mappings that preserve collinearity and ratios of distances. Besides translations and rotations, affine deformations can also scale and shear the object [70]. In 3D space, affine deformations can be represented with a single square matrix with 12 parameters (3 rotations, 3 translations, 3 scaling factors and 3 shearing factors) and can be written as:

$$M_{aff} = Sh(\gamma_{xy}, \gamma_{xz}, \gamma_{yz}) Sc(s_x, s_y, s_z) M_{rig}$$

where:

$$Sh(\gamma_{xy}, \gamma_{xz}, \gamma_{yz}) = \begin{pmatrix} 1 & \tan \gamma_{xy} & \tan \gamma_{xz} & 0 \\ 0 & 1 & \tan \gamma_{yz} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad Sc(s_x, s_y, s_z) = \begin{pmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

with $\frac{\pi}{2} - \gamma_{xy}$, $\frac{\pi}{2} - \gamma_{xz}$ and $\frac{\pi}{2} - \gamma_{yz}$ being the angles between the coordinate pairs after the shearing transformation is applied, and s_x, s_y, s_z the scaling factors along each axis. Some example rigid and affine 2D transformations can be seen in Figure 2.2.

Non-rigid transformations

Non-rigid transformations are mappings where every voxel in the image can be transformed independently. In most medical image registration applications, these transformations have to be bijective, thus ensuring a one-to-one mapping between the two images. Breaking this criteria is called ‘folding’ and it results in information loss and broken topology. Non-rigid models can be *non-parametric* or *parametric*.

In **non-parametric approaches**, the deformation field is directly optimised through the registration process. These models transform each voxel independently and thus need regularisation to constrain the solution space to plausible deformations. In the following paragraphs a few different approaches on how this is achieved are presented.

The **optical flow** method [89] optimises the SSD between a fixed and a moving image by computing its first derivative. In order to constrain the deformation field, the sum of the Laplacian of the deformation field is added as a penalty term [89]. This method is efficient, but not very robust and is dependent on finding a good weighting factor between the smoothness regulariser and the dissimilarity measure.

The **Demons** method [93] has been introduced by Thirion *et al.* in 1998. In this framework the deformation field is calculated through an iterative process which updates the field with a normalised optical flow between the two images (see Figure 2.5).

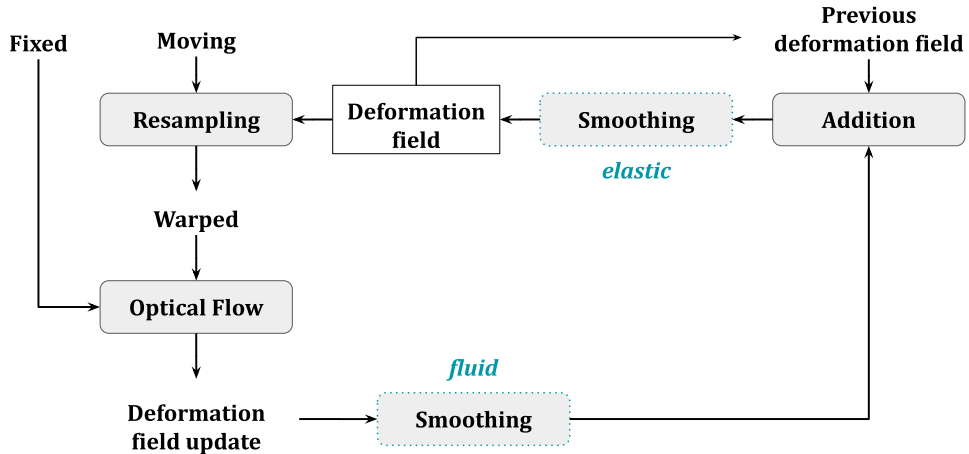


Figure 2.5: Illustration of the Demons algorithm [93] showing the input images (fixed and moving), and the warped image obtained after resampling with the current iteration of the deformation field. When smoothing is applied to the deformation field update ($\delta \mathbf{u}$) it is called *fluid-like* regularisation, while when smoothing is applied after the update ($\mathbf{u}^t + \delta \mathbf{u}$) it is called *elastic-like* regularisation.

Regularisation of the predicted deformation field is incorporated through Gaussian smoothing, either before or after the deformation field update. When done before, the approach is called *fluid-like regularisation*:

$$\mathbf{u}^{t+1} = \mathbf{u}^t + \mathbf{G} * \delta \mathbf{u} \quad (2.15)$$

while when done after the update, it is called *elastic-like regularisation*:

$$\mathbf{u}^{t+1} = \mathbf{G} * (\mathbf{u}^t + \delta \mathbf{u}) \quad (2.16)$$

where \mathbf{G} is a Gaussian kernel, $*$ is the convolution operator, and $t + 1$ is the current iteration. When $t = 0$, $\mathbf{u}^t = Id$ (the identity grid).

Beg *et al.* [94] introduced the **non-stationary velocity field** model, also known as the **large deformations diffeomorphic metric mapping (LDDMM)** framework, in which time-varying velocity fields are integrated over time to generate a deformation field. Diffeomorphisms are one-to-one smooth and continuous mappings which are also invertible (non-zero Jacobian determinant) [75], *i.e.*, they preserve topology. Composing two diffeomorphisms results in a diffeomorphism, meaning that you can compose many deformations and still have an inverse.

In LDDMM, the deformation field is defined through the following ordinary differential equation (ODE):

$$\frac{\partial \varphi(\mathbf{x}, t)}{\partial t} = \mathbf{v}(\varphi(\mathbf{x}, t), t) \quad (2.17)$$

with initial condition $\varphi(\mathbf{x}, 0) = Id$ and where the solution at $t = 1$:

$$\varphi(\mathbf{x}, 1) = \int_0^1 \mathbf{v}(\varphi(\mathbf{x}, \tau), \tau) d\tau \quad (2.18)$$

is diffeomorphic.

In their method, Beg *et al.* [94] discretised the time-varying field into a number of steps which are then composed to generate the final deformation field.

The use of a single **stationary velocity field (SVF)** instead of the time-varying one was simultaneously proposed by Ashburner *et al.* [62] and Hernandez *et al.* [95, 96]. For SVFs, the ODE is defined as:

$$\frac{\partial \varphi(\mathbf{x}, t)}{\partial t} = \mathbf{v}(\varphi(\mathbf{x}, t)) \quad (2.19)$$

with initial condition $\varphi(\mathbf{x}, 0) = Id$ and where the solution at $t = 1$ is given by:

$$\varphi(\mathbf{x}, 1) = \int_0^1 \mathbf{v}(\varphi(\mathbf{x}, \tau)) d\tau \triangleq \exp(\mathbf{v}) \quad (2.20)$$

The advantage of using stationary instead of time-varying velocity fields is brought by the concept of *scaling and squaring* [66], a method which makes the integration much faster. More specifically, the velocity field can be divided into $n = 2^i$ steps, and the final deformation field φ can be calculated through Euler integration starting from the identity grid in $\log_2 n$ steps. For example, for $n = 8$, the deformation field is computed in four steps:

$$\begin{aligned} \varphi^{1/8} &= Id + \mathbf{v}/8 \\ \varphi^{2/8} &= \varphi^{1/8} \circ \varphi^{1/8} \\ \varphi^{4/8} &= \varphi^{2/8} \circ \varphi^{2/8} \\ \varphi^{8/8} &= \varphi^{4/8} \circ \varphi^{4/8} \end{aligned} \quad (2.21)$$

Moreover, the inverse transformation $\varphi^{-8/8}$ can be easily computed through backward integration, by starting from $\varphi^{-1/8} = Id - \mathbf{v}/8$.

Vercauteren *et al.* [97] introduced a diffeomorphic version of the Demons algorithm [93]. Using properties of Lie group theory, their implementation ensures a one-to-one mapping between the reference and floating images through composition (instead of addition) of diffeomorphic transformations, starting from the identity grid. Their method uses a *scaling and squaring* approach to compute the vector field exponentials which allows for an efficient implementation when compared to other approaches. Similar to the original Demons [93], Vercauteren *et al.* [97] describe their proposed diffeomorphic algorithm for both *fluid-like* and *elastic-like* regularisation strategies. In this case, equation 2.15 (*fluid*) becomes:

$$\varphi^{t+1} = \varphi^t \circ \exp(\mathbf{G} * \mathbf{v}) \quad (2.22)$$

and 2.16 (*elastic*) becomes:

$$\varphi^{t+1} = \mathbf{G} * (\varphi^t \circ \exp(\mathbf{v})) \quad (2.23)$$

Christensen *et al.* [98] introduced an **inverse consistent approach** in order to ensure a one-to-one mapping between the fixed and moving images. A forward transformation $\varphi_0 : \Omega_F \rightarrow \Omega_M$ is optimised at the same time as the backward transformation $\varphi_1 : \Omega_M \rightarrow \Omega_F$, and the optimisation process takes into account two similarity measures and a constraint on the inverse consistency:

$$\begin{aligned} \mathcal{L}(\mathbf{F}, \mathbf{M}(\varphi)) &= \mathcal{D}_{\text{SSD}}(\mathbf{F}(\varphi_1), \mathbf{M}) + \mathcal{D}_{\text{SSD}}(\mathbf{F}, \mathbf{M}(\varphi_0)) \\ &+ \underbrace{\sum \|\varphi_0(\mathbf{x}) - \varphi_1^{-1}(\mathbf{x})\|^2 + \|\varphi_1(\mathbf{x}) - \varphi_0^{-1}(\mathbf{x})\|^2}_{\text{inverse consistency constraint}} \end{aligned} \quad (2.24)$$

This penalisation makes sure that the optimised fields (φ_0, φ_1) are as close as possible to their respective approximated inverses $(\varphi_1^{-1}, \varphi_0^{-1})$.

Beg *et al.* [99] introduce inverse consistency in their proposed LDDMM framework in two ways: the *consistent-integral-cost* which evaluates the registration match between the two images at all time points, and the *consistent-midpoint-cost* which evaluates the match at $t = \frac{1}{2}$. Similarly, Avants *et al.* [60] proposed an approach where they estimated transformations to a middle space between the two images. In their case, the *halfway* forward transformation and the inverse of the *halfway* backward transformation are composed in order to calculate the image similarity between the deformed moving image and the fixed image.

Parametric approaches are a different class of non-rigid registration methods. Unlike non-parametric approaches, they rely on a function to generate the deformation field. In this case, the number of model parameters is lower than the number of voxels, but smoothness constraints are still used in order to favour continuous transformations. In the following paragraphs, four image registration parametric approaches are presented.

The **spatial normalisation using basis functions** approach was introduced by Ashburner *et al.* [75] and it uses a linear combination of discrete cosine transform basis functions to describe the spatial transform. This algorithm optimises the parameters of a deformation field which minimises the SSD between a warped and a fixed image. To penalise folding, the authors also introduce penalty terms to the cost function.

Shen *et al.* [100] introduced the **hierarchical attribute matching mechanism for elastic registration algorithm** where attribute vectors are used for each voxel to drive the registration process. More specifically, these vectors include information about the voxel's underlying tissues and its neighbourhood. Such information is obtained through performing segmentation of the images. To make the registration algorithm execute faster, not all voxels are used, but are mostly selected from the tissue boundaries. One major drawback of this approach is the dependence on the quality of the pre-processing steps and the existence of a large number of parameters that need to be set to achieve good performance.

Rueckert *et al.* [68] introduced the **free-form deformation (FFD)** algorithm, a method which is based on cubic B-Spline interpolation. In their approach, a mesh of control points is used to parametrise the deformation. When a control point is moved, the local neighbourhood moves as well. As the support of the basis functions spans across 4 control points, the area of the moving neighbourhood becomes $(4\delta_x \cdot 4\delta_y \cdot 4\delta_z)$ voxels around that respective control point, where δ_i represents the spacing between 2 adjacent control points along the i^{th} axis in the lattice. The choice of δ and the number of control points is defined by the image size and the spacing along each axis.

To compute the new coordinate of a point, the following formula is used:

$$\mathbf{T}(\vec{x}) = \sum_{l=0}^3 \sum_{m=0}^3 \sum_{n=0}^3 B_l(u)B_m(v)B_n(w)\mu_{i+l,j+m,k+n} \quad (2.25)$$

where:

$$\begin{aligned} u &= \frac{\mathbf{x}}{\delta_x} - \lfloor \frac{\mathbf{x}}{\delta_x} \rfloor \\ v &= \frac{\mathbf{y}}{\delta_y} - \lfloor \frac{\mathbf{y}}{\delta_y} \rfloor \\ w &= \frac{\mathbf{z}}{\delta_z} - \lfloor \frac{\mathbf{z}}{\delta_z} \rfloor \end{aligned}$$

are the relative positions of the index point, i, j, k are the indices of the first control point to be taken into account, B_l, B_m, B_n are the approximated third-order spline polynomials applied along each axis and μ_i, μ_j, μ_k are the first control point positions. In this framework, Rueckert *et al.* [68] introduced the bending energy penalty term to ensure smoothness of the deformation field. This, however, does not guarantee a one-to-one mapping between the reference and the floating image, and others [92, 91] constrained it further with the use of Jacobian determinant penalties (see Section 2.1.2).

Later, Ruckert *et al.* [101] introduced the **parametric stationary velocity field** method, which is a diffeomorphic version of the FFD approach, through composition of control point grids. The interpolation scheme used, as well as the choice of NMI as a similarity measure, made this algorithm computationally expensive. Modat *et al.* [102] introduced a SVF diffeomorphic and symmetric registration model, where the velocity field was parameterised with cubic B-spline basis functions. The proposed method used NMI as a similarity measure and introduced a regularisation term based on the Jacobian determinant of the deformation field (see Section 2.1.2).

2.1.4 Diffusion tensor image registration

Image registration of DT-MRI can enable better alignment of WM tracts than what is possible when using structural MRI data only [61, 103]. As an alternative, scalar-valued data can be used instead of the higher-order tensors, and, in fact, FA maps are a popular choice in many neuroimaging studies [104] as they highlight the main WM tracts. In the scalar case, an image transformation guided by a deformation field φ simply changes the location of each point x , mapping x to $\varphi(x)$. When dealing with higher-order data, however, further steps need to be applied. This subsection summarizes the challenges of registering DT-MR images brought forward by the orientational information contained by this data.

Tensor reorientation. For DTI images, interpolation is not as straightforward as scalar-valued data, and Figure 2.6 illustrates this schematically. In panel A, an anisotropic region of an axial slice of a DT image is shown, while the other two panels showcase what happens when this region is rotated anti-clockwise around the z-axis with 30° . In panel B, the tensor components are simply interpolated at the warped locations, but this procedure does not preserve the original internal organisation of the region. In panel C, the individual tensors are also rotated around the z-axis with 30° , thus retaining the local anatomy.

Tensor reorientation was initially described by Alexander *et al.* [105]. One algorithm for this procedure is called the *finite strain* strategy where the reorientation matrix can be computed at each point in the deformation field φ through a polar decomposition of the local Jacobian matrix. Through this factorisation $J = RP$, the non-singular matrix J is split into a unitary matrix R (representing the rotation), and a positive-semidefinite Hermitian matrix P [106]. The resulting rotation matrix R can then be used to reorient the diffusion tensors.

Image similarity measures based on tensor differences. As for every other image registration algorithm, an appropriate image similarity needs to be designed. The most popular DTI ‘dissimilarity’ measure is the **euclidean distance squared** [61, 108]. Assuming $\mathbf{M}(\varphi(\mathbf{x}))$ is the deformed moving image with reoriented diffusion

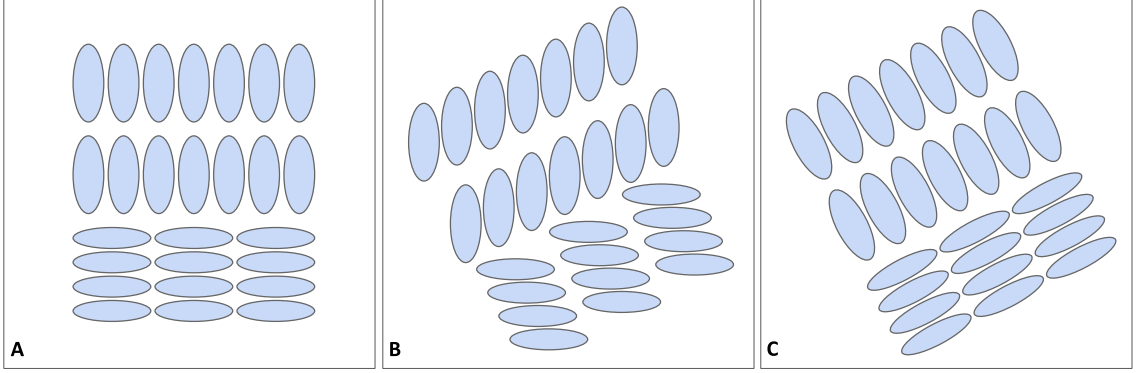


Figure 2.6: DTI reorientation showing: an anisotropic region in A, the same region after rotating it anti-clockwise around the z-axis with 30° in B, and the correct tensor interpolation which preserves the original internal organisation of this region in C. Image adapted from Zhang *et al.* [107].

tensors, the **euclidean distance squared** is defined as:

$$\mathcal{D}_{\text{EDS}}(\mathbf{F}, \mathbf{M}(\varphi)) = \sum_{\mathbf{x} \in \Omega_{\mathbf{F}}} \|\mathbf{F}(\mathbf{x}) - \mathbf{M}(\varphi(\mathbf{x}))\|_C^2 \quad (2.26)$$

where the **euclidean distance** between **two tensors** \mathbf{D}_1 and \mathbf{D}_2 is:

$$\|\mathbf{D}_1 - \mathbf{D}_2\|_C = \sqrt{\text{Tr}((\mathbf{D}_1 - \mathbf{D}_2)^2)} \quad (2.27)$$

Zhang *et al.* [61] also propose the **euclidean distance squared between deviatoric tensors** as a measure which is less sensitive to the isotropic components of diffusion tensors. It is defined as:

$$\mathcal{D}_{\text{DDS}}(\mathbf{F}, \mathbf{M}(\varphi)) = \sum_{\mathbf{x} \in \Omega_{\mathbf{F}}} \|\mathbf{F}(\mathbf{x}) - \mathbf{M}(\varphi(\mathbf{x}))\|_D^2 \quad (2.28)$$

where the *deviatoric*, D , of diffusion tensor \mathbf{D} is equal to: $\mathbf{D} - \frac{1}{3}\text{Tr}(\mathbf{D})\mathbf{I}$ and \mathbf{I} is the identity tensor [108]. Therefore, the **euclidean distance between two deviatoric tensors** \mathbf{D}_1 and \mathbf{D}_2 is defined as:

$$\|\mathbf{D}_1 - \mathbf{D}_2\|_D = \sqrt{\frac{8\pi}{15} \left(\|\mathbf{D}_1 - \mathbf{D}_2\|_C^2 - \frac{1}{3}\text{Tr}^2(\mathbf{D}_1 - \mathbf{D}_2) \right)} \quad (2.29)$$

Image registration frameworks for diffusion MRI. Some of the original registration methods [109, 110] did not take tensor reorientation into account, thus introducing errors in matching the images. Zhang *et al.* [61] introduced a piecewise affine algorithm known today as *DTI-TK*, where the novelty came from both the explicit optimisation of the tensors reorientation and from their proposed derivative-based formulation. Moreover, the piecewise affine transformations were merged together to generate a smooth deformation field.

Cao *et al.* [111] developed an LDDMM framework for registration of diffusion tensors, by matching their corresponding principal eigenvectors. Later, Yeo *et al.* [112] extended the diffeomorphic version of the Demons algorithm [97] to work with tensor data. In *DTI-Demons*, they introduced a derivation of the exact finite strain differential and showed that using the exact gradient led to better registration results. Modat *et al.* [113] extended *Nifty-Reg* to work with diffusion tensor data as well. The short list presented here of tensor-based registration algorithms is not exhaustive and a more complete review of DT-based image registration frameworks can be found in [114, 115].

Besides the popular rank-2 diffusion tensor, other higher-order diffusion data can be used in image registration applications. This is because DTI cannot model crossing fibers [116, 117]. More advanced diffusion imaging methods, such as the HARDI acquisition protocol [117], can be used to better characterise regions with crossing fibre populations. For example, the spherical deconvolution technique allows the direct estimation of the distribution of fiber orientations within each voxel from diffusion weighted (DW)-MRI data [118]. Raffelt *et al.* [119] propose the use of ODF data for image registration purposes and extend the ANTs symmetric diffeomorphic normalisation method [60] to work with them. Moreover, Uus *et al.* [56] propose the use of a similarity metric based on angular correlation [120], instead of the original SSD one.

2.2 Medical image segmentation

Segmentation is the process of delineating an image into regions of interest (ROIs) based on their color, gray level, texture, or contrast. In medical imaging, this division has the additional property of classifying these ROIs based on their anatomical function, or in order to separate normal from abnormal tissue. For example, segmenting an MR image of the brain could consist of separating WM from cGM and deep gray matter (dGM), or identifying the spread of a tumour from the surrounding healthy organ. The resulting segmentation maps can then be used in upstream analysis, such as measuring the volume of different brain tissue types during the neonatal developmental stages, or during neurodegenerative disease progression.

Automatic image segmentation is not a trivial task as more often than not medical images contain artifacts, and can suffer from partial volume effects or intensity inhomogeneities. Many algorithms have been proposed for medical image segmentation, but due to its complex and challenging nature it still remains an active area of research [121, 122]. For the purpose of this thesis, the focus will be on medical image segmentation of brain MRI, with a particular interest in neonatal data. The following sections will present the most common medical image segmentation techniques, grouped into: registration-based approaches and intensity-based approaches. Other methods such as level sets and active shape/appearance models [123, 124, 125, 126] are out of scope for this thesis.

2.2.1 Registration-based approaches

Some of the most promising methods for segmenting brain MRI use image registration as one of their steps. These methods rely on the existence of an atlas (prior knowledge of brain morphology), together with its corresponding label maps, to produce segmentation maps of a subject's brain image. More specifically, an image registration algorithm determines how to warp the brain atlas into the to-be-segmented image space, after which the atlas labels are propagated using the predicted deformation. The following subsections will present different approaches to this technique, grouped by whether they use a single atlas approach, a multiple atlas approach, or a probabilistic atlas approach.

Single atlas approaches

The first atlas-based brain segmentation methods were using a single atlas to predict labels for unseen subjects [64, 127]. As illustrated in Figure 2.7, the manual segmentation of a single brain was propagated onto a given subject in order to find its corresponding tissue labels.

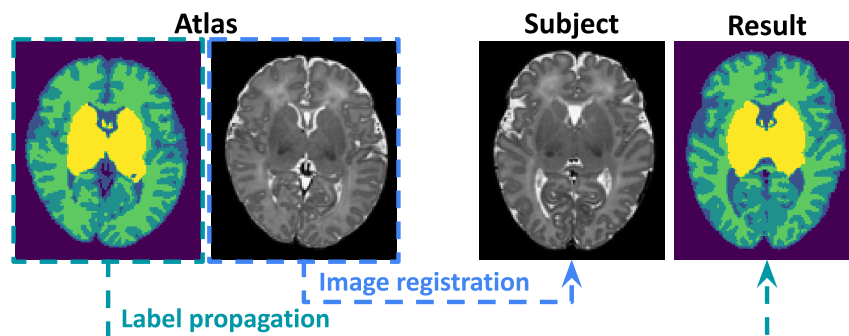


Figure 2.7: Illustration of medical image segmentation based on a single atlas registration approach. An image registration method is first used to align the template and the subject. Then, the atlas labels are mapped to the coordinates of the to-be-segmented image.

This technique is rarely used today due to its reliance on a single anatomical instance without taking into account the variability found across human brains. Moreover, it relies on accurate image registration between the two images in order to propagate the reference labels. This is not always possible, as individual variability in cortical folding prevents accurate inter-subject anatomical correspondences. Moreover, in early development, there are additional time-dependent changes which occur both in tissue microstructure (due to, for example, the ongoing myelination which affects the MRI contrast) and anatomy (such as cortical folding).

Multi-atlas approaches

The use of multiple atlases achieved better results than using a single anatomy [91, 128, 129]. This is shown schematically in Figure 2.8 where multiple atlases are registered to a subject, and the resulting deformation fields are used to propagate the reference labels. As a final step, the segmentation maps are fused into one using averaging, non-uniform weighting [130] or atlas selection [131]. In fact, besides the accuracy of the registration method and the quality of the atlases themselves, the performance of multi-atlas approaches also relies on the fusion process. For this reason, besides majority voting, other label fusion methods have also been proposed [132, 133, 134], but they are out of scope for this thesis.

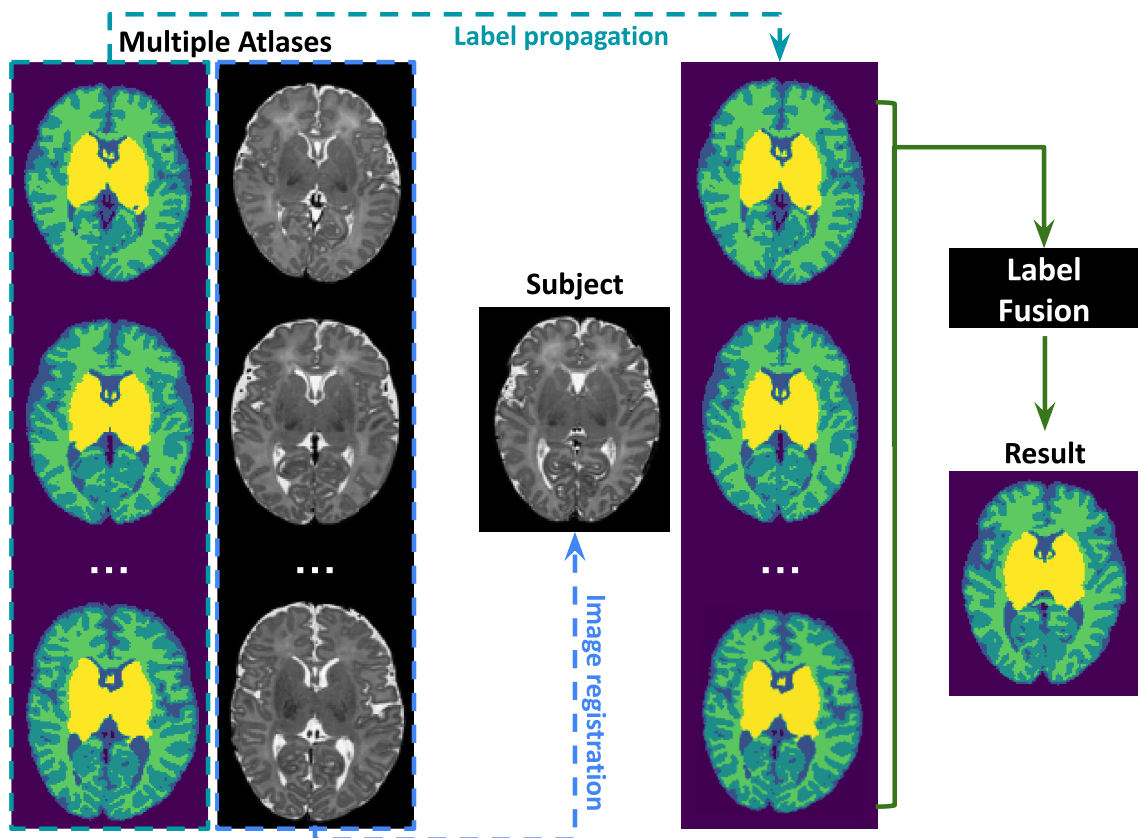


Figure 2.8: Illustration of medical image segmentation based on the multi atlas registration approach. In this case, multiple templates are registered with a subject, and the resulting deformation fields are used to propagate the atlas labels. The resulting segmentation is achieved through *label fusion*, a method which aggregates the multiple label maps into one.

Probabilistic atlas approaches

Probabilistic atlases are created by combining brain images of a representative cohort, and can show variation over populations and/or time. For this, a large enough dataset of segmented brain images is needed to capture the anatomical variability of

the studied population. Probabilistic atlas approaches for segmenting brain images rely on a statistical model of image intensities together with *a priori* knowledge of different brain tissues (*i.e.*, the probabilistic atlas). This is shown schematically in Figure 2.9 where tissue probability maps are warped onto the subject space to infer this prior knowledge, which is then used in an expectation maximisation (EM)-type scheme to produce the final result.

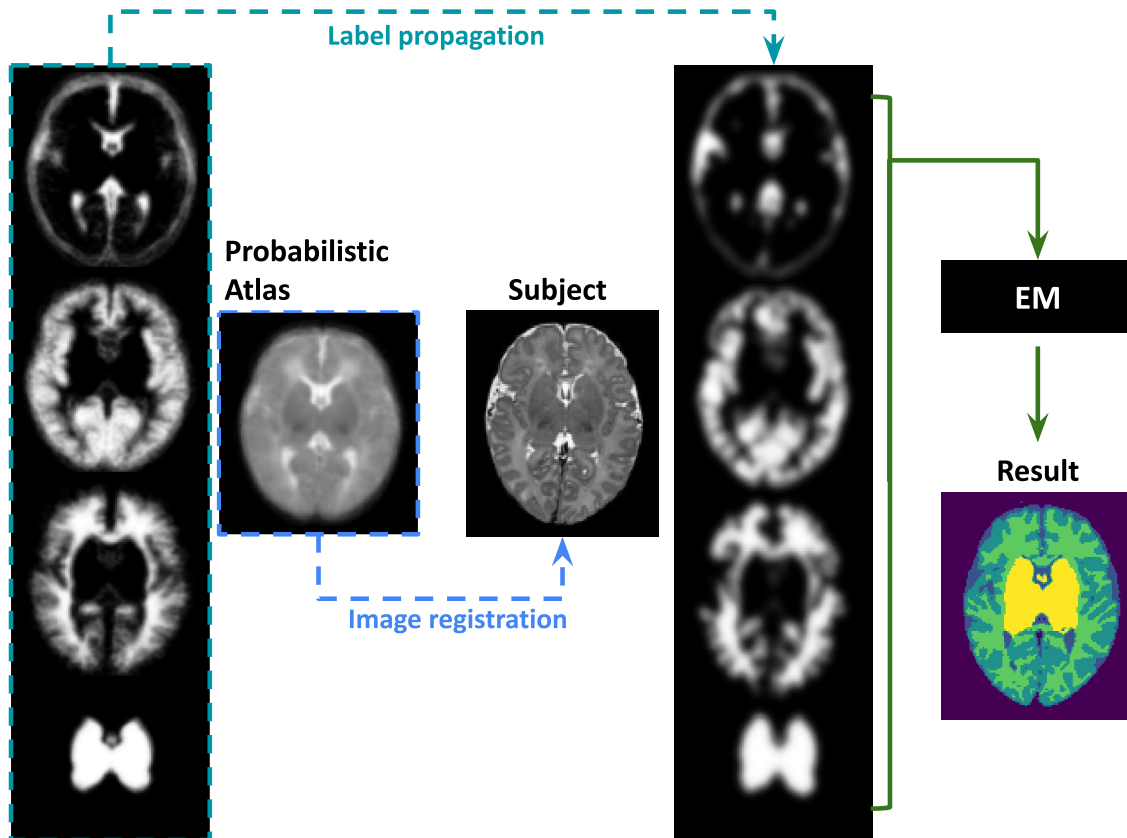


Figure 2.9: Illustration of medical image segmentation based on the probabilistic atlas registration approach. In this case, a probabilistic atlas is registered with a subject, and the resulting deformation fields are used to propagate the tissue probability maps. The resulting probabilistic labels then become the *a priori* knowledge used in medical image segmentation algorithms based on statistical models of intensities, offering information about the spatial distribution of different brain tissues.

The rapid development of the neonatal brain brings unique challenges to the segmentation problem. For this reason, the use of a spatio-temporal probabilistic atlas, which consists of multiple age-specific atlases, is often required in order to capture the anatomical variability of this cohort [12]. Other methods have also tried relaxing the prior label probabilities by iteratively adapting the atlas to the subject's anatomy [135, 136, 137].

Finally, one potential drawback of these approaches comes from the use of average brain templates which have blurrier boundaries when compared to an individual atlas. This means that the image registration step can often generate less accurate alignments between the probabilistic atlas and the to-be-segmented subject.

2.2.2 Intensity-based approaches

Intensity-based image segmentation approaches are methods which classify each voxel in an MR image based on its intensity, and are grounded in computer vision standard classifiers such as Gaussian mixture models (GMM), K-Means or K-nearest neighbours (K-NN). Automatic segmentation of brain MRI is not an easy task, and most algorithms require the following preprocessing steps in order to achieve the desired results:

- **Brain extraction**, also known as skull-stripping, is a procedure through which the skull and non-brain tissues, such as neck and fat, are removed from brain MRI scans. This is often a crucial step as these regions may have overlapping intensity distributions with the to-be-segmented brain tissue.

In the medical image community, one of the most widely used algorithms is the brain extraction tool, or BET, [138] which is also part of the FSL software package [139]. BET is a physics-based model which uses a closed surface that evolves to fit the brain. Alternatively, a multi-atlas-based skull stripping approach [140] can be used, in which predefined templates with corresponding brain masks are aligned with the to-be-segmented subject. Then, the labels are propagated onto the subject space, and fused to create the final brain mask. More recently, with the advent of medical image deep learning, neural networks have been trained to perform skull stripping [141]. The main advantage is that, once an algorithm has been trained, the inference time on a new and unseen MR image is very fast. On the other hand, deep learning techniques are not yet generalisable to all types of MR datasets, vendors and subject biases.

- **Bias field correction** is needed as spatial intensity inhomogeneities are often observed in MR images. These non-uniformities exist as a result of magnetic field variations and present themselves as a smooth low-frequency spatially varying intensity change which affects the MR image [142].

Algorithms which try to solve this problem are often applied as a pre-processing step (such as the N3 bias correction framework [143], or its improved variant, N4 [144]). Alternatively, some image segmentation frameworks interleave the segmentation process with bias field correction, achieving both at the same time [145].

- **Motion correction** is also an important preprocessing step as motion is ubiquitous in MRI because the time required for the majority of MR sequences to collect the necessary data is much longer than most types of physiological motion, including respiratory motion, vessel pulsation, CSF flow and even involuntary patient motion. At best, bulk motion can lead to slice misalignment which can be corrected for with registration algorithms [146]. However, if motion happens during the acquisition part of the experiment, it can lead to blurring of object edges, ghosting, loss of information or undesired strong signals [146].

In MRI acquisition of the neonatal brain there are 2 types of motion that can typically happen: rigid motion, caused by head movements, or non-rigid motion, caused by arterial pulsation or other internal sources [146]. Even though neonatal brain MRI protocols are shorter than adult scans, infants cannot be prevented to move and motion artefacts will still be present in the acquired images [147]. One potential solution is to sedate the infants, which is the case with some of the babies scanned for the ePrime dataset [35] for which parents gave consent, but this is not the preferred procedure and the goal is to image unsedated neonates [148]. In dHCP [11], for example, infants were scanned during natural sleep without sedation, where there is a risk of sporadic movement. For this reason, motion correction techniques of the acquired MRI acquisition, such as those proposed by Cordero-Grande *et al.* [149, 39] and validated on neonatal datasets, are therefore needed to obtain motion-free images.

The remainder of this section will present K-NN clustering, K-Means clustering and Gaussian Mixture models. Finally, we will describe the EM algorithm on which the state-of-the-art brain segmentation frameworks usually rely.

K-nearest neighbours

K-NN is a non-parametric supervised method used for both classification and regression. As it is a supervised method, it requires training data, such as pairs of features and their corresponding labels. In case of classification tasks, the algorithm outputs a class membership for each queried data point. This is achieved through plurality vote: an object is classified as class c if the majority of its closest K neighbours (from the training data) are part of class c .

Figure 2.10 shows an example of the algorithm for $K = 7$, where the training data is shown in panel (a) with their corresponding labels (red, green or blue), together with 2 data points to be classified (the black diamonds). Panel (b) of the figure shows, for each object, its 7 nearest neighbours based on the most widely used distance measure (Euclidean distance). Finally, the last panel shows the resulting class for each object.

This algorithm can be applied to image segmentation problems and, in fact, it was adapted to work with adult and neonatal brain MR images by Warfield *et al.* [150]. In their case, in addition to using the image intensities as features, they enhance the classification process through using an anatomical template to moderate the segmentation.

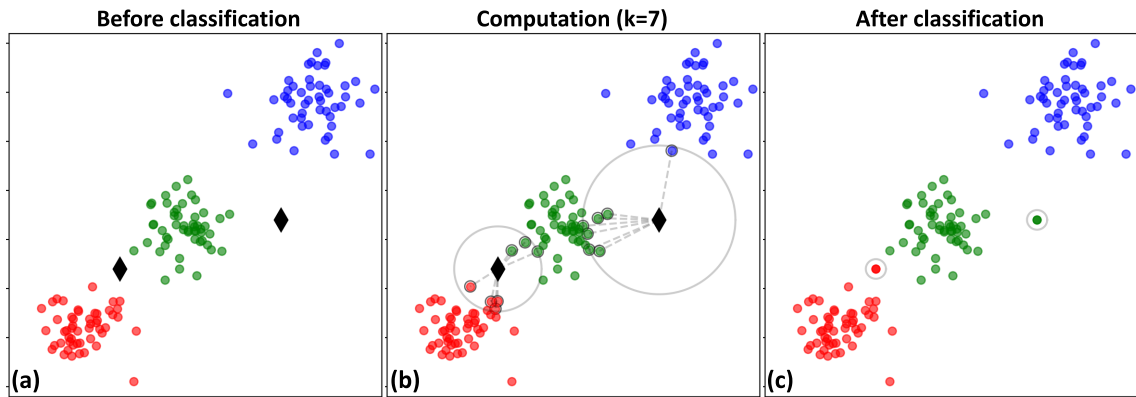


Figure 2.10: K-nearest neighbours example for $K = 7$. In (a) the training data consisting of 3 distinct classes (red, green and blue) is shown together with 2 objects which are to be classified (the black diamonds). K of their respective nearest neighbours are highlighted in (b), while (c) shows the final classification result.

K-Means clustering

K-Means is an unsupervised clustering algorithm which, as the name suggests, aims to divide an image into K clusters such that each observation (or voxel in the image) belongs to the cluster with the nearest mean (center/centroid). More specifically, K-Means clustering minimizes within-cluster variances. Mathematically, it aims to optimize the objective function:

$$\sum_{i=1}^n \sum_{k=1}^K \|y_i - v_k\|^2 \quad (2.30)$$

where Y is the image to be segmented consisting of n voxels with intensities (y_1, y_2, \dots, y_n) , and v_k is the centroid of the k^{th} cluster.

The number of clusters K is a hyper-parameter which needs to be decided beforehand. A poor choice of K can often yield unwanted results and it is therefore important to run a diagnostic check in order to choose the appropriate value. The steps involved in the algorithm are:

1. Choose value for K
2. Randomly initialise centroids
3. Calculate cluster membership for each data point
4. Re-calculate centroids based on the assigned data points
5. Repeat steps 3 and 4 until no (or a small) change in the centroids is observed.

Figure 2.11 shows an example of the algorithm for $K = 3$, where the random initialisation of the centroids and the initial datapoint assignment to each cluster are

shown in Figure 2.11(a)–(b), while the iterative part of the above algorithm (steps 3–4) is shown in panels (c)–(e). Last panel (f) shows the final result.

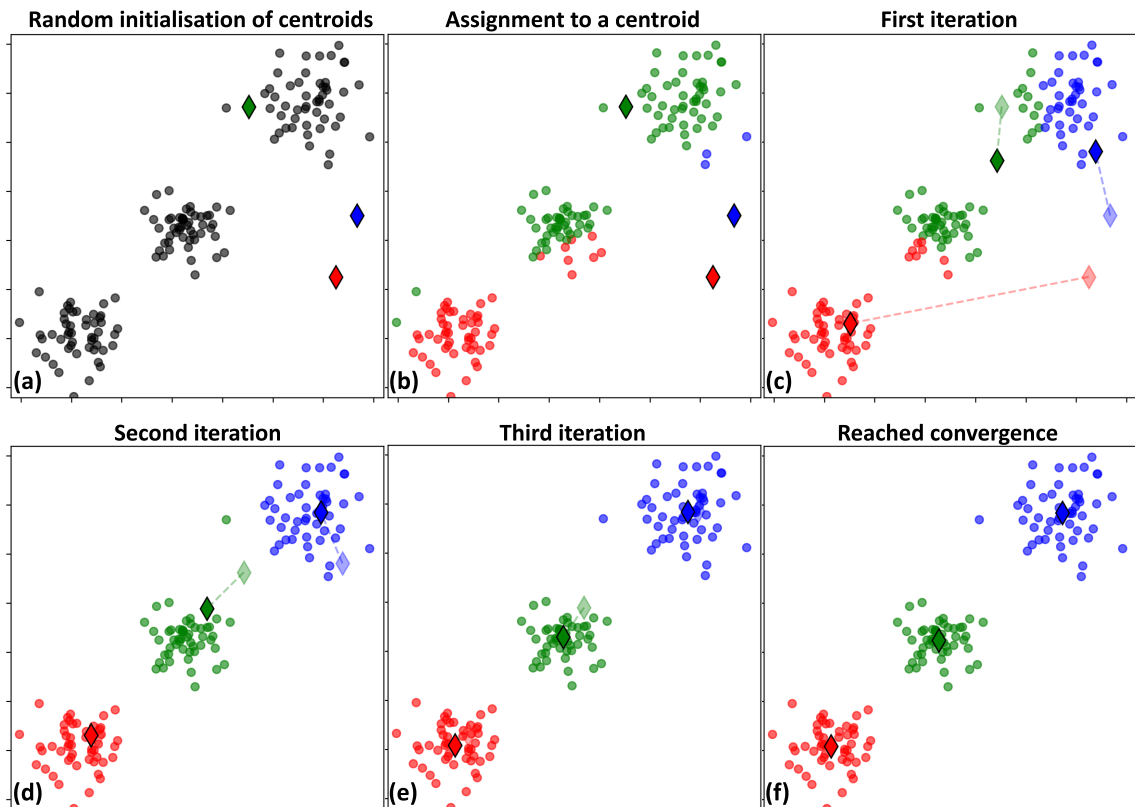


Figure 2.11: K-Means clustering example for $K = 3$, where the dataset (black circles) to be classified together with the random initialisation of the clusters' centroids (red, blue and green diamonds) are shown in (a). In (b) each data point has been assigned to a cluster based on the distance to the centroids. (c), (d) and (e) show the first three iterations of the algorithm as it re-calculates the centroids and reassigns the datapoints to the 3 clusters. Finally, in (f), after a few iterations, the algorithm has converged and the centroids do not change anymore.

Although K-Means has been successful in medical image segmentation applications [151], it is susceptible to noise and outliers, as well as the initial choice of centroids. **Fuzzy C-means clustering** [152] is a soft version of K-Means, where each data point has a probability of belonging to each cluster, instead of exclusively belonging to one class only. In fact, fuzzy C-means is a generalised version of K-Means which introduces membership values w_{ik} (the degree to which data point i belongs to class k) with $\sum_{k=1}^K w_{ik} = 1$. This is a helpful property for medical image segmentation due to partial volume effects [153].

Gaussian mixture models

A GMM is a probabilistic model for representing normally distributed clusters within a dataset. More specifically, it attempts to find a mixture of multi-dimensional Gaussian probability distributions that best model the input data. A GMM with

K components is parameterized by three variables: each k^{th} component's mean μ_k and variance/covariance σ_k , and mixing coefficients ω_k (where $\sum_{k=1}^K \omega_k = 1$). Such a model can be used to perform image segmentation by describing the likelihood of a brain MRI voxel as belonging to a tissue class. The probability of observing intensity y is described by:

$$p(y) = \sum_{k=1}^K \omega_k \mathcal{N}(y|\mu_k, \sigma_k) \quad (2.31)$$

where:

$$\mathcal{N}(y|\mu_k, \sigma_k) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(-\frac{(y - \mu_k)^2}{2\sigma_k^2}\right) \quad (2.32)$$

When K is known or set to a reasonable value, the most commonly used method to solve the mixture of Gaussians is the EM algorithm [154].

Expectation-Maximization

EM is a technique for maximum likelihood estimation generally used when there exists a closed form expression for updating the model parameters. Moreover, it is an iterative algorithm which is guaranteed to approach a local maximum (or saddle point). EM for mixture models starts with an initialization step which assigns model parameters to reasonable values based on the data, and then alternates between two steps until convergence:

- The first step, or the **initialization step**, consists of randomly (or based on some initial estimated values) assigning values for the GMM components means $\hat{\mu}_k$, variances $\hat{\sigma}_k$, and mixing coefficients $\hat{\omega}_k$.
- **The E-step (expectation step)** consists of calculating the expectation of component k for each data point $y_i \in Y$ (the probability that y_i is generated by component C_k) given the estimated model parameters $\hat{\mu}_k$, $\hat{\sigma}_k$, and $\hat{\omega}_k$:

$$\hat{p}_{ik} = \frac{\hat{\omega}_k \mathcal{N}(y_i|\hat{\mu}_k, \hat{\sigma}_k)}{\sum_{j=1}^K \hat{\omega}_j \mathcal{N}(y_i|\hat{\mu}_j, \hat{\sigma}_j)}$$

- **The M-step (maximization step)** will then update the current parameter estimation by maximizing the expectations calculated in the E-step:

$$\begin{aligned} \hat{\omega}_k &= \frac{1}{n} \sum_{i=1}^n \hat{p}_{ik} \\ \hat{\mu}_k &= \frac{\sum_{i=1}^n \hat{p}_{ik} y_i}{\sum_{i=1}^n \hat{p}_{ik}} \\ \hat{\sigma}_k^2 &= \frac{\sum_{i=1}^n \hat{p}_{ik} (y_i - \hat{\mu}_k)^2}{\sum_{i=1}^n \hat{p}_{ik}} \end{aligned}$$

where n is the number of data points / voxels.

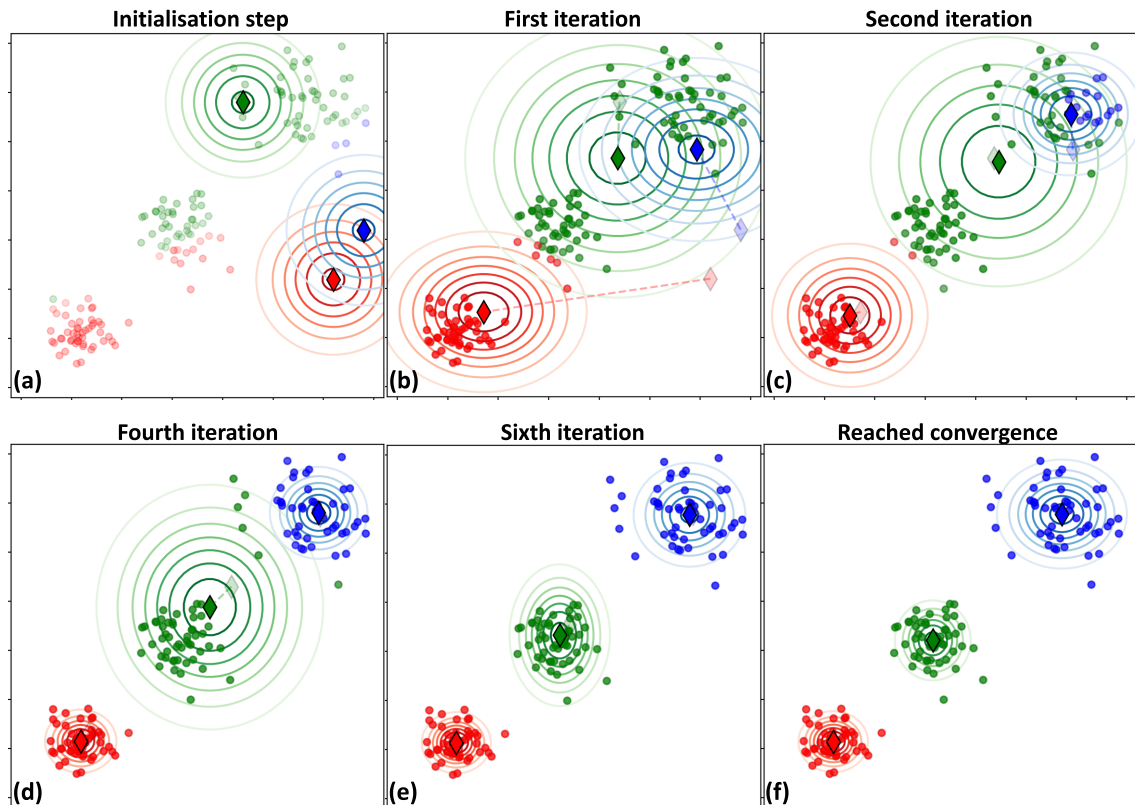


Figure 2.12: Expectation-Maximization clustering example for $K = 3$, where the initialisation step of the Gaussians (red, blue and green) are shown in (a). In (b), (c), (d) and (e) the first, second, fourth and sixth iterations of the algorithm are shown as it alternates between the E-step and the M-step. Finally, in (f), the algorithm has converged and the calculated parameters do not change anymore.

Figure 2.12 shows an example of the algorithm for $K = 3$, where the parameter initialisation and subsequent datapoint assignment to each cluster (based on the highest probability) is shown in Figure 2.12(a). The iterative part of the above algorithm is shown in panels (b)–(e), while the last panel (f) shows the final result, when the model has reached convergence.

The EM algorithm has been successfully applied to brain segmentation since Wells *et al.* [155]. In this case, the **E-step** estimates the soft segmentation of a brain MRI given the current estimate of the model parameters, while the **M-step** estimates the parameters for the intensity distribution of each tissue class. In addition, it can be extended to include partial volume estimation [156, 157], bias field correction [158, 145], registration parameters [145], spatially constraining anatomical priors for the tissue class probabilities, as well as neighbourhood statistics by means of a Markov random field (MRF) [159, 160]. The latter was used by Habas *et al.* [161] to segment the fetal brain into: skull, CSF, GM, WM, germinal matrix and ventricles. Moreover, as strong anatomical priors could negatively bias the segmentation, Cardoso *et al.* [135, 136, 137] introduced a relaxation of the prior tissue probabilities for infant brain segmentation with structural abnormalities. Melbourne *et al.* [162] extended this approach to contain outlier rejection of intensity clusters which have large Mahalanobis distance from the predicted model.

2.2.3 Medical image segmentation frameworks

This section presents the most common segmentation frameworks used in the medical imaging community, with a focus on EM-based software packages as they are one of the most widely used brain segmentation approaches today, while the deep learning-based segmentation models will be presented in Section 3.2.2. A more comprehensive review on neonatal brain image segmentation methods can be found in [163], as well as the 2012 MICCAI Grand Challenge on Neonatal Brain Segmentation (NeoBrainS12) which has been summarised in [164].

Ashburner *et al.* [145] developed the popular image analysis software framework, **statistical parametric mapping (SPM)**, which contains, among others, tools for adult brain image segmentation. More specifically, their approach is EM-based, which includes segmentation and bias correction, as well as non-rigid registration of a probabilistic atlas. The latter improves the accuracy of the predicted segmentation maps and enhances the robustness of the bias correction step. Wang *et al.* [165] used SPM for the NeoBrainS12 challenge [164] in conjunction with the probabilistic atlas developed by Kuklisova-Murgasova *et al.* [12]. As a post-processing step, they used connected component analysis to correct partial volume errors.

Another state-of-the-art medical image registration and segmentation method is the **advanced normalization tools (ANTs)** software package [60, 166]. More specifically, Atropos [166] performs image segmentation and uses the EM framework with contextual information by means of MRF. In fact, ANTs was used by Wu *et al.* [167] for the NeoBrainS12 challenge, together with N4 for bias correction [144], and SyN [60] for atlas registration.

Cardoso *et al.* [137] introduce **NiftySeg** as an EM-based framework with a prior relaxation strategy [168] which aims to iteratively adapt the probabilistic atlas to the anatomy of the subject such that images with high anatomical variability can also be segmented. For the NeoBrainS12 challenge, Melbourne *et al.* [162] extended this approach with an outlier strategy which rejected intensity clusters that were too far away from the predicted model.

Finally, Makropoulos *et al.* [169, 170] introduce the **developing region annotation with expectation maximisation (Draw-EM)** algorithm for automatic multi-atlas neonatal brain image segmentation. Similarly to others, their method also includes bias field correction [144], partial volume correction and spatial regularisation [170]. More specifically, instead of parcellating the brain into different tissue classes (such as WM, GM and CSF) based on intensities only, the authors use *a priori* information extracted from the ALBERTs [171] dataset to segment the neonatal brain into 87 sub-cortical, cortical and cerebral structures. This prior knowledge dataset consists of manually segmented 18 sub-cortical and 32 cerebral structures in MR images of 5 healthy term-born neonates and 15 preterm-born babies, imaged between 36.6 weeks and 44.9 weeks PMA [171].

The algorithm starts by performing brain extraction [138] and N4 bias field correction [144] on each to-be-segmented subject. Then, the ALBERTs images are rigidly, affinely, and non-rigidly registered to the subject space, and the resulting deformation fields are used to propagate their respective labels. A locally-weighted label fusion scheme is then used to generate the atlas priors. Subject-specific tissue prior probabilities are also created through K-Means [172] (with $K = 4$, for WM, GM, CSF, and extra-cranial space) and an initial simple EM scheme. The tissue and atlas priors are then used to subdivide the brain into 87 structures, as the initial labelling of the ALBERTs dataset included clusters which contained both WM and GM intensities into one label. One of the assumptions of the EM scheme is that every region follows a Gaussian distribution, which, in the case of the initial ALBERTs division, was not true. Moreover, as some of the structures will have similar intensities (*e.g.*, different cortical gray matter sub-structures), Draw-EM introduces a hierarchical mixture model where the same Gaussian distribution is shared amongst such regions. This EM scheme is run until convergence, and the final output is made up of 87 structures or 9 tissue labels [148]. Figure 2.13 summarizes the Draw-EM steps.

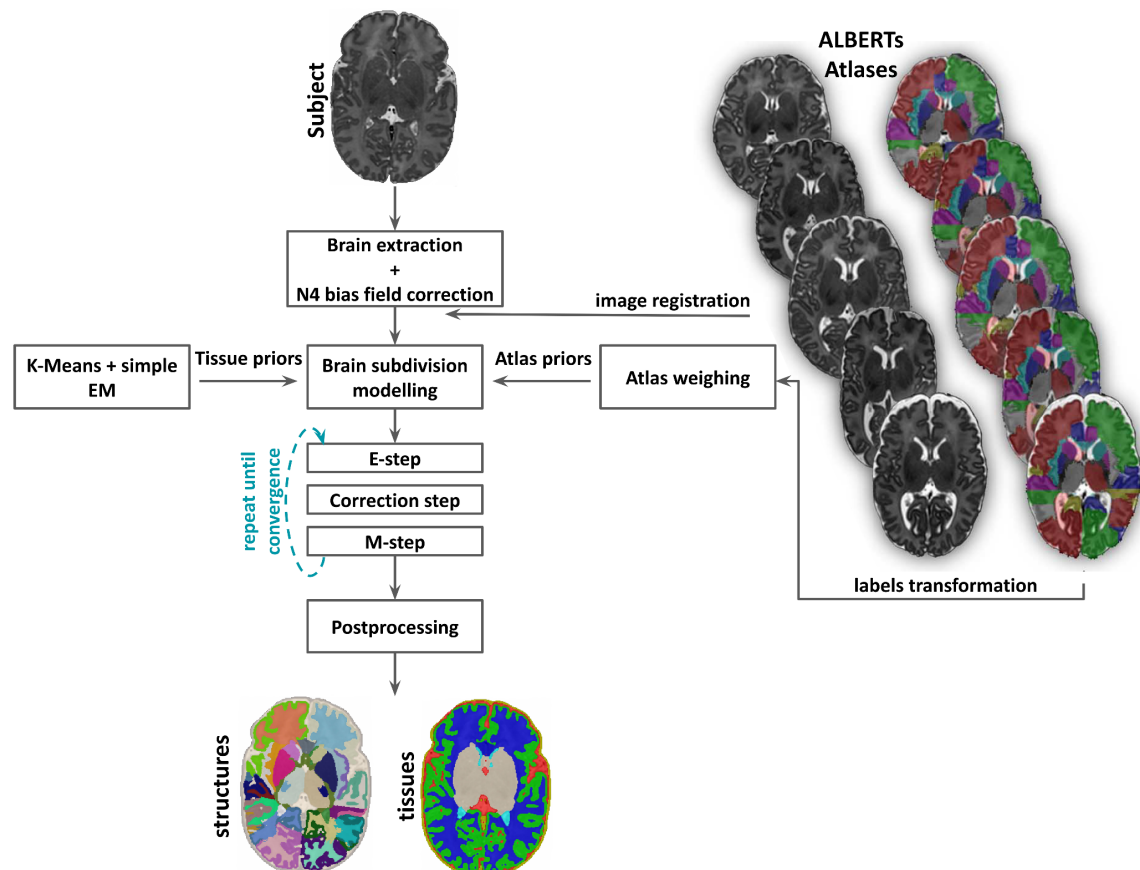


Figure 2.13: Draw-EM pipeline showing how the ALBERTs dataset [171] is used to predict segmentation maps (structures and tissues) for an unseen neonatal subject. Image adapted from Makropoulos *et al.* [170].

This method was evaluated on a large cohort of 234 mainly preterm infants and

showed good results. It is part of the MIRTk¹ package and it was used in this thesis as the main software toolbox to generate tissue labels for the dHCP and ePrime cohorts used for the studies.

2.2.4 Medical image segmentation evaluation

An important step of any image segmentation algorithm is measuring its performance against ground truth labels or other models. This section presents the most common metrics for validating image segmentation models, with a focus on spatial overlap and surface based metrics.

Let a medical image be represented by a set of points $X = \{x_1, x_2, \dots, x_n\}$, where $|X| = w \times h \times d = n$ represents the cardinality of the set X , and w, h, d are the volume's width, height and depth, respectively. Let the ground truth label map be defined as $S_g = \{S_g^b, S_g^f\}$, and the predicted segmentation as $S_p = \{S_p^b, S_p^f\}$, such that $S_g^b(x) + S_g^f(x) = 1$, $S_p^b(x) + S_p^f(x) = 1$, and $S_g^{b|f}(x) \in [0, 1]$, $S_p^{b|f}(x) \in [0, 1]$, for $\forall x \in X$. The b and f superscripts represent the background and the foreground (anatomy of interest) classes, respectively. For the sake of simplicity, but without loss of generality, the rest of this subsection will focus on binary segmentations only.

Spatial overlap metrics

There are 4 basic measures that reflect the overlap between the two volumes and they are called: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) [173]. Figure 2.14 shows an example ground truth and predicted segmentations, together with the aforementioned measures. The TP is equal to the number of correctly predicted foreground class voxels:

$$TP = |S_g^f \cap S_p^f| \quad (2.33)$$

The TN is the number of correctly predicted background class voxels:

$$TN = |S_g^b \cap S_p^b| \quad (2.34)$$

The FP is the number of predicted foreground voxels which should have been background:

$$FP = |S_g^b \cap S_p^f| \quad (2.35)$$

Finally, the FN is the number of predicted background voxels which should have been foreground:

$$FN = |S_g^f \cap S_p^b| \quad (2.36)$$

¹<https://github.com/BioMedIA/MIRTk>

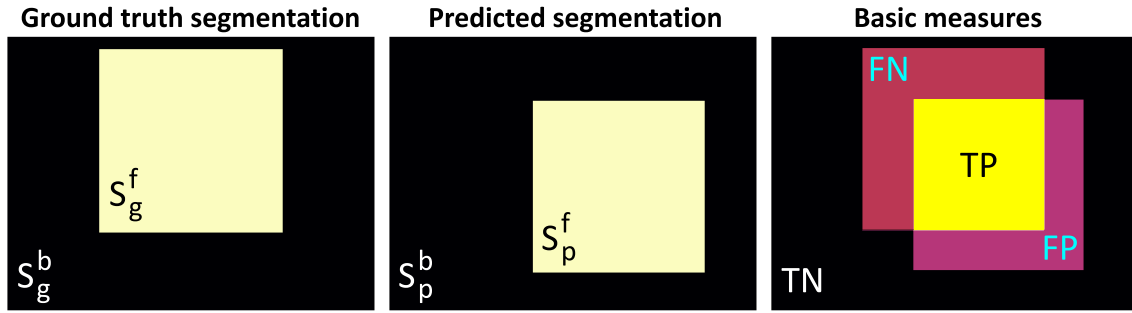


Figure 2.14: Schematic example of a ground truth binary segmentation (background class S_g^b and foreground class S_g^f) and a predicted segmentation (background class S_p^b and foreground class S_p^f), together with the four basic measures: TP (the intersection between the foreground classes of both ground truth and predicted segmentations), TN (the intersection between the background classes of the ground truth and predicted segmentations), FP (the incorrectly predicted foreground class) and FN (the incorrectly predicted background class).

These four measures can be represented in a confusion matrix, as seen in Figure 2.15. Moreover, all spatial overlap metrics can be derived from these four basic measures.

		Prediction			
		foreground	background		
Ground truth	foreground	TP (correct hit)	FN (underestimation)	Recall $TP / (TP+FN)$	FNR $FN / (TP+FN)$
	background	FP (overestimation)	TN (correct rejection)		
		Precision $TP / (TP+FP)$	FOR $FN / (TN+FN)$	Specificity $TN / (FP+TN)$	Fallout $FP / (FP+TN)$

Figure 2.15: Confusion matrix showing the TP, TN, FP and FN measures, as well as 6 commonly derived metrics: FNR, recall, fallout, specificity, precision and FOR.

The two most important metrics are **precision** and **recall** [173]. The **positive predicted value (PPV)**, also known as **precision**, is the fraction of relevant instances among the retrieved instances. It is defined as:

$$PPV = \text{Precision} = \frac{TP}{TP + FP} \quad (2.37)$$

and it is a good metric to quantify over-segmentation in the predicted labels.

The **true positive rate (TPR)**, also known as **sensitivity** or **recall**, is defined as the fraction of correctly identified foreground class voxels in the ground truth:

$$\text{TPR} = \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2.38)$$

and, as a complementary measure to precision, it is often used to quantify under-segmentation.

There are other metrics that can be derived from the four basic measures, such as the true negative rate (TNR), the false positive rate (FPR) and the false negative rate (FNR) [173]. TNR, also known as specificity, calculates the fraction of correctly classified background class voxels in the ground truth background segmentation:

$$\text{TNR} = \text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (2.39)$$

The FPR, also called fallout, measures the fraction of incorrectly predicted voxels in the ground truth background segmentation:

$$\text{FPR} = \text{Fallout} = \frac{\text{FP}}{\text{TN} + \text{FP}} = 1 - \text{TNR} \quad (2.40)$$

while the FNR is defined as:

$$\text{FNR} = \frac{\text{FN}}{\text{TP} + \text{FN}} = 1 - \text{TPR} \quad (2.41)$$

Equations 2.38 and 2.41, as well as equations 2.39 and 2.40, are equivalent to each other and for this reason it is not common to report both of them together for validation purposes. For the sake of completion, the false omission rate (FOR) was introduced in Figure 2.15, but in practice it is not used for validating image segmentation models.

The most prevalent metric in the literature is the **Dice score coefficient (DSC)**, also known as the **F1 score** (or the harmonic mean of the precision and recall), and, for the foreground class, it is defined as [174]:

$$\text{DSC} = \frac{2|S_g^f \cap S_p^f|}{|S_g^f| + |S_p^f|} = \frac{2 \text{TP}}{2 \text{TP} + \text{FP} + \text{FN}} \quad (2.42)$$

The **Jaccard index (JAC)** is also an overlap metric and it is defined as the intersection between the predicted and ground truth class over their union:

$$\text{JAC} = \frac{|S_g^f \cap S_p^f|}{|S_g^f \cup S_p^f|} \quad (2.43)$$

However, these two metrics are related:

$$\text{JAC} = \frac{|S_g^f \cap S_p^f|}{|S_g^f \cup S_p^f|} = \frac{2|S_g^f \cap S_p^f|}{2(|S_g^f| + |S_p^f| - |S_g^f \cap S_p^f|)} = \frac{\text{DSC}}{2 - \text{DSC}} \quad (2.44)$$

which means that, in practice, just one of them should be reported as a validation metric.

Surface-based metrics

Spatial distance metrics are used to evaluate image segmentation tasks especially where the boundary (contour) is of importance. Let X and Y be two finite point sets, then one can define the **directed Hausdorff distance (HD)** as:

$$\vec{d}_H(X, Y) = \max_{x \in X} \min_{y \in Y} d(x, y) \quad (2.45)$$

where $d(x, y)$ is a measure of distance between the two points, *e.g.* Euclidean distance. In practice, the undirected **HD** is used and this is the maximum between the two directed HD distances:

$$d_H = \max\{\vec{d}_H(X, Y), \vec{d}_H(Y, X)\} \quad (2.46)$$

This metric, however, can be sensitive to noise and outliers. It is therefore recommended to use the q^{th} quantile of distances instead of the maximum, where q is most commonly chosen as 95 [175].

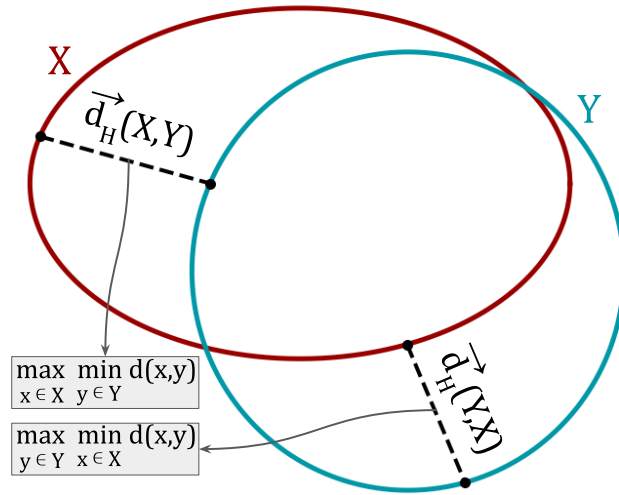


Figure 2.16: Schematic representation of the directed Hausdorff distance between two sets of points X and Y .

Finally, the **average surface distance (ASD)**, also known as the average Hausdorff distance, is the HD averaged over all points:

$$d_{AH} = \max\{\vec{d}_{AH}(X, Y), \vec{d}_{AH}(Y, X)\} \quad (2.47)$$

where $\vec{d}_{AH}(X, Y)$ is the **directed ASD** given by:

$$\vec{d}_{AH} = \frac{1}{|X|} \sum_{x \in X} \min_{y \in Y} d(x, y) \quad (2.48)$$

ASD is often used as an image segmentation validation measure as it is less sensitive to outliers than HD. Other metrics exist, such as the Mahalanobis distance, but they are out of scope for this thesis. A more detailed investigation of medical image segmentation metrics can be found in the review paper by Taha *et al.* [173].

Deep learning for medical image analysis

This chapter offers an overview of deep learning medical image analysis methods, starting with an introduction into the field in **Section 3.1**, and followed by state-of-the-art techniques in **Section 3.2**.

Section 3.1 presents the main concepts important for understanding deep learning methods, with a focus on convolutional neural networks (**Section 3.1.1**), training a neural network (**Section 3.1.2**), and an overview of neural network architectures (**Section 3.1.3**) and training strategies (**Section 3.1.4**).

Section 3.2 delves into the more advanced techniques, focusing on deep learning image registration and segmentation techniques (**Sections 3.2.1** and **3.2.2**), as well as visual attention (**Section 3.2.3**) and domain adaptation (**Section 3.2.4**), respectively.

3.1 Deep learning theory

Deep learning is a subset of machine learning and artificial intelligence which uses neural networks with multiple processing layers to learn hierarchical representations of data [176]. There are three main types of deep learning methods: supervised, unsupervised and deep reinforcement learning, although the latter is out of scope for this thesis. Since its recent ground-breaking success in computer vision and speech recognition, deep learning has become a dominant trend in medical image analysis [177]. More specifically, deep learning has been successfully used for medical image segmentation and classification [178, 179, 180], image registration [181, 182, 183], image fusion [184], computer-aided diagnosis [185], and lesion detection [186], among others.

This section focuses on the theory behind the artificial neural network (ANN), and the convolutional neural network (CNN), describes best practices for training these models, and presents the most common neural network architectures, with a focus on the ones important for this thesis.

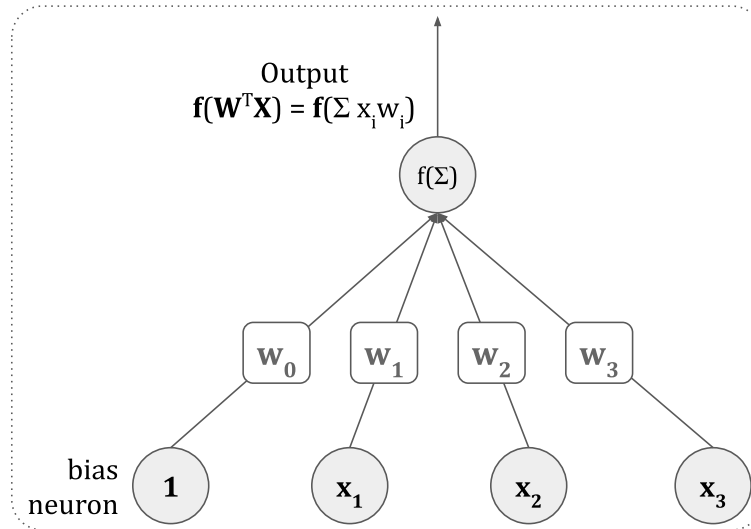
3.1.1 Artificial neural networks

The main building block for any deep learning framework is the perceptron (see Figure 3.1a), introduced in 1958 by Frank Rosenblatt [187]. When stacking multiple perceptrons together, as shown in Figure 3.1b, we arrive at the general architecture for any artificial neural network, the multilayer perceptron (MLP).

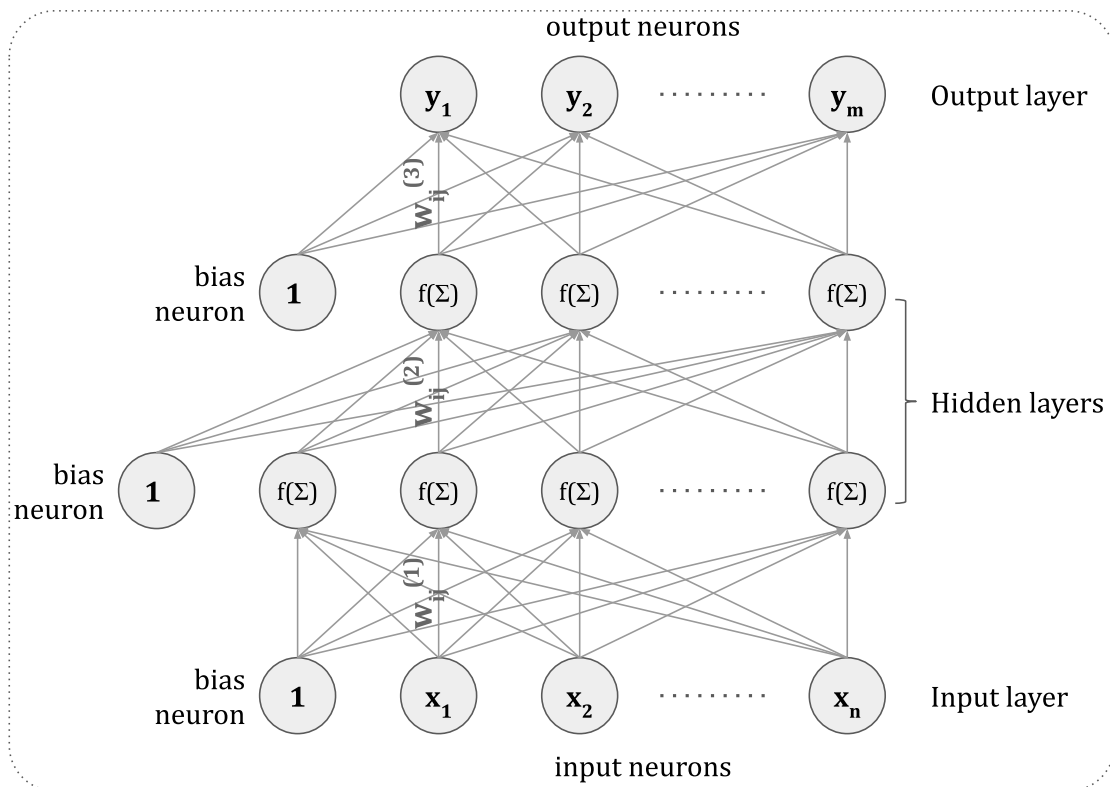
MLPs are made of an input layer, one or more hidden layers and one output layer. Connections between pairs of neurons from adjacent layers have a weight w attached to them, signifying the *strength* of those connections. Information flows from the input layer to the output layer, with the input neurons relaying the data as is (without modifying it), while the hidden layers' neurons apply an activation function f to the weighted sum of incoming values:

$$\sum_i w_i x_i$$

. A network with full connectivity between any two adjacent layers is called a fully connected network.



(a) The structure of an artificial neuron is composed of the inputs x_i (and the bias), the weights w_i , and the output $f(\sum_i x_i w_i)$, where f is called an activation function.



(b) The structure of a multilayer perceptron with one input layer, 2 hidden layers and one output layer.

Figure 3.1: Main building blocks of artificial neural networks showing in (a) the perceptron and in (b) a multilayer perceptron.

Convolutional neural networks

The most popular type of neural network for analysing images is the CNN. A CNN is very similar to a fully connected network, with the added constraint that not all of the pixels or voxels in the input image are connected to the neurons from the convolutional layers. In fact, only the voxels that fall under the *receptive field* of a certain neuron are connected to that neuron, as shown in Figure 3.2.

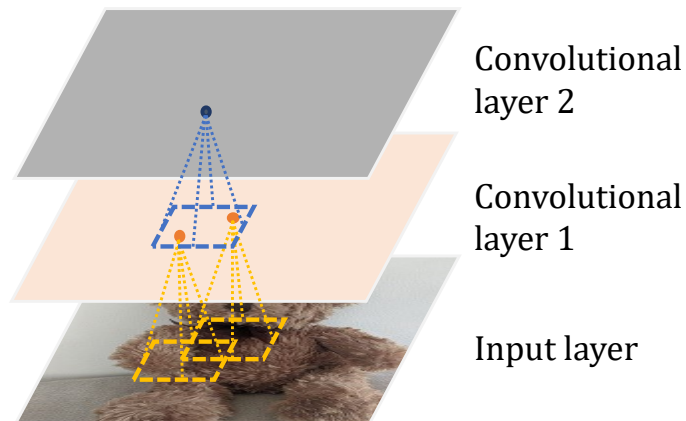


Figure 3.2: Schematic representation of two CNN layers together with their local receptive fields. Image adapted from Geron *et al.* [188].

Convolutional layer. Convolutional layers are the most important building blocks of a CNN. A two-dimensional convolutional layer receives an input object of dimensions $\text{Width}_1 \times \text{Height}_1 \times \text{Channels}_1$, and outputs an object of dimensions $\text{Width}_2 \times \text{Height}_2 \times \text{Channels}_2$. The connection between the two sets of variables determining the input size and the output size comes from the following equations:

$$\begin{aligned}\text{Width}_2 &= (\text{Width}_1 - F + 2P)/S + 1 \\ \text{Height}_2 &= (\text{Height}_1 - F + 2P)/S + 1 \\ \text{Channels}_2 &= K\end{aligned}$$

where the four hyperparameters are: K - the number of filters, F - the spatial extent of filters, S - the stride, and P - the amount of zero padding. A 2D convolutional layer introduces $F^2 \cdot \text{Channels}_1$ weights for each filter $k \in K$. Figure 3.3 shows a 3×3 filter sliding over the input image which was zero-padded, with stride 1 in panel a, and stride 2 in panel b. Figure 3.3 c shows a 3×3 filter with dilation rate 2 which is also known as *atrous convolution* [189]. Dilation represents the spacing between the values in a kernel, and it covers a wider field of view at the same computational cost (*i.e.*, a 3×3 filter with dilation rate 2 covers the same area as a 5×5 kernel with dilation rate 1).

Convolutional layers are used to extract features from the input images. It is generally thought that the first convolutional layer focuses on low level features of the image, while the following layers assemble the previous layers' features into higher-order representations [188].

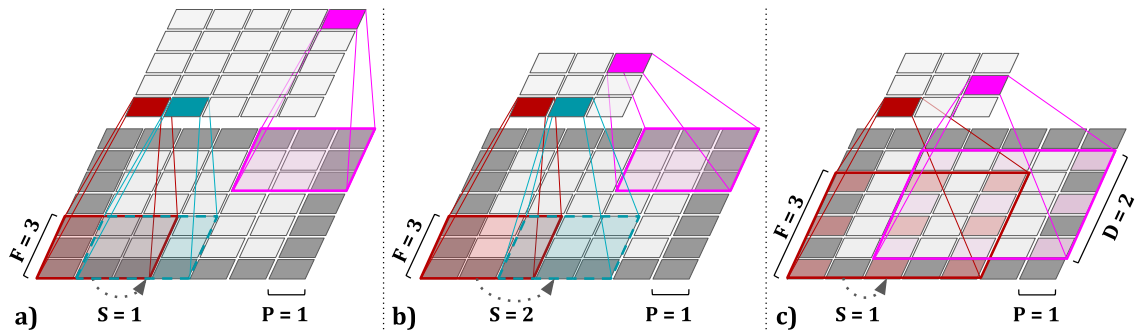


Figure 3.3: Example of convolutional layers with filters of size 3×3 ($F = 3$), and zero-padding of 1, with: a) stride $S = 1$, b) stride $S = 2$, and c) stride $S = 1$ and dilation rate $D = 2$.

Pooling layer. Another common layer found in a CNN is the pooling layer. The most widely used one is called *max-pooling*, followed closely by the *average-pooling* layer. A two-dimensional pooling layer receives an input volume of $\text{Width}_1 \times \text{Height}_1 \times \text{Channels}_1$, and outputs a volume of $\text{Width}_2 \times \text{Height}_2 \times \text{Channels}_2$, with the following properties:

$$\begin{aligned}\text{Width}_2 &= (\text{Width}_1 - F)/S + 1 \\ \text{Height}_2 &= (\text{Height}_1 - F)/S + 1 \\ \text{Channels}_2 &= \text{Channels}_1\end{aligned}$$

It introduces zero weights as it computes a fixed function of the input (either the maximum value in the receptive field or the average value, respectively). Moreover, it is uncommon to use padding in these layers. Figure 3.4 shows an example of a max-pooling layer and an average-pooling layer.

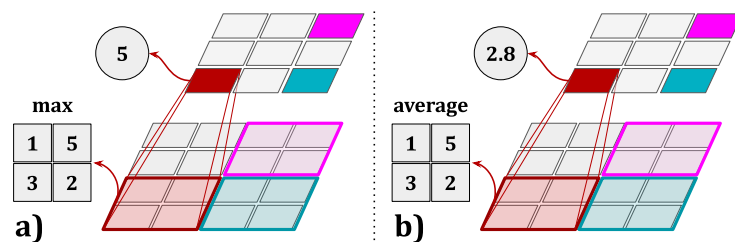


Figure 3.4: Example of: a) max-pooling and b) an average-pooling layer with filters of size 2×2 , and stride 2.

3.1.2 Training a neural network

To train an ANN, Rumelhart *et al.* [190] introduced the *backpropagation* algorithm. In short, this algorithm requires two stages: a *forward pass* and a *backward pass*. In the forward pass, data samples are fed to the network which passes them through each layer, computing the weighted sums and activations, and returns the output values. Then, a *loss function* is calculated representing the error between the current prediction of the model and the desired output. Working in reverse, the error

gradient for all connections is then calculated until the input layer is reached. Finally, *gradient descent* is used to tweak the weights of the connections in order to reduce the loss. For this algorithm to work, the activation functions used throughout the network should be continuous, differentiable or well-defined [191] so that error gradients can be computed.

Activation functions. Thus, the first activation function that became popular is the sigmoid function:

$$\sigma(z) = \frac{1}{1 + \exp(-z)} \quad (3.1)$$

as it is continuous and differentiable at every point of its domain. Another popular activation function is the hyperbolic tangent:

$$\tanh(z) = \frac{2}{e^z + e^{-z}} \quad (3.2)$$

which, unlike the sigmoid, can have negative output values (see Figure 3.5).

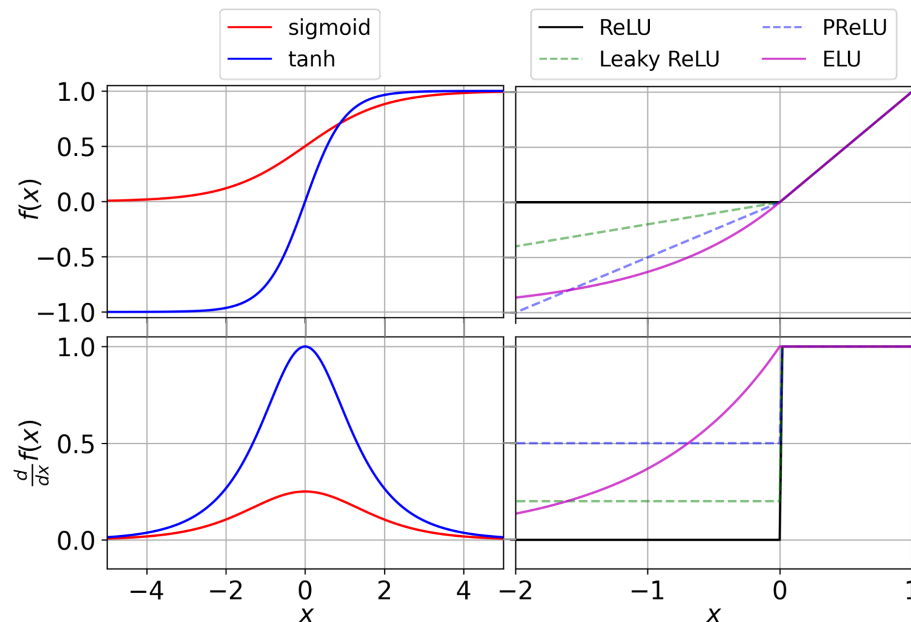


Figure 3.5: Example of six of the most popular activation functions (first row), as well as their corresponding derivatives (second row). The first column shows the sigmoid and the hyperbolic tangent, while the second column shows the ReLU function and three of its variants: Leaky ReLU with $a = 0.2$, PReLU with $\alpha = 0.5$, and ELU with $a = 1.0$.

Finally, the most widely used activation functions today are the Rectified Linear Unit (ReLU) function:

$$\text{ReLU}(z) = \max(0, z) \quad (3.3)$$

and its different variants: LeakyReLU, Parametric Rectified Linear Unit (PReLU),

and Exponential Linear Unit (ELU) [192]:

$$\begin{aligned}
 \text{LeakyReLU}(z) &= \max(0, z) + a \min(0, z) \\
 \text{PReLU}(z) &= \max(0, z) + \alpha \min(0, z) \\
 \text{ELU}(z) &= \begin{cases} z, & \text{if } z > 0 \\ a(e^z - 1), & \text{if } z \leq 0 \end{cases}
 \end{aligned}
 \tag{3.4}$$

Here, α is a parameter which is learnt during training, while a is a hyperparameter to be chosen. These four activation functions are shown in Figure 3.5, together with their derivatives. It is important to mention that, even though ReLU-based functions are not differentiable when $z = 0$, sub-gradients can be used and optimised with gradient descent [193, 194].

When a neural network is trained for a classification task, such as determining whether an image represents a ‘cat’ or a ‘dog’ (exclusive classes), the output layer will have a *softmax* activation function. This function transforms the predicted values, also known as *logits*, into probabilities, such that the output of each neuron in the final layer will correspond to the estimated probability of the class. For class $k \in K$ total number of classes, *softmax* is written as:

$$\text{softmax}(\mathbf{z})_k = \frac{\exp(z_k)}{\sum_{j=1}^K \exp(z_j)}
 \tag{3.5}$$

where \mathbf{z} is a vector containing the logits of the output layer the network, as seen Figure 3.6.

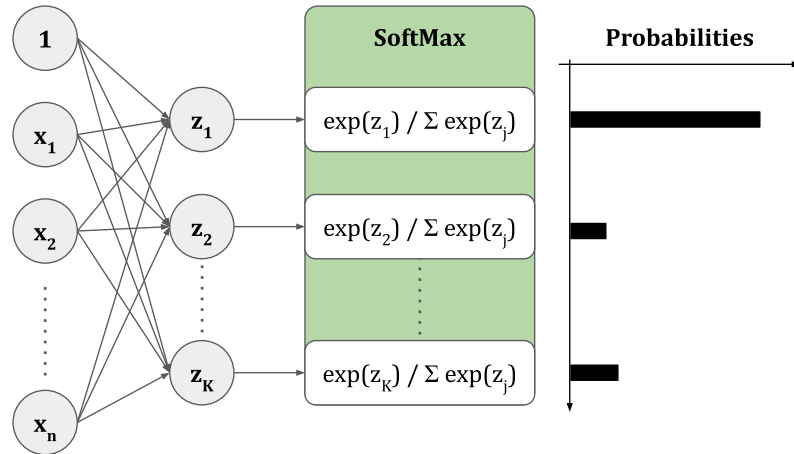


Figure 3.6: Example of the *softmax* activation function applied on the logits of a layer, thus transforming the values into exclusive probabilities for each of the K classes.

Vanishing/Exploding gradients. A common problem found in training deep neural networks is called the *vanishing gradients* problem. This happens when the gradients computed through backpropagation become small enough that the gradient descent step does not change the weights significantly or at all, and so the training cannot converge. In fact, networks using the sigmoid and the hyperbolic tangent functions often suffer from this problem due to their derivative becoming very small when $\sigma(z)$ approaches 0 or 1, and when $\tanh(z)$ approaches -1 or 1. At the opposing pole, the *exploding gradients* problem happens when very large error gradients accumulate and the algorithm diverges. This problem happens mostly in recurrent neural networks, which are out of scope for this thesis. Glorot and Bengio [195] showed that by employing a certain initialisation strategy for the neural network’s weights, the vanishing/exploding gradients problem can be alleviated. This solution is called *Xavier initialization* when using the sigmoid activation function, and *He initialization* when using the ReLU activation function.

Another proposed solution which makes the network more stable while training is called *batch normalisation* [196]. As the name suggests, the operation is performed on the batches axis (see Figure 3.7). Batch normalisation is often added before applying the activation function of a layer. It zero-centres and normalises, then scales and shifts the inputs to that layer. The scaling and shifting factors are two per-layer parameters which are learnt by the network. As the concept of ‘batch’ is not always present, especially in medical image applications where training is often performed with a batch size of 1 or 2 [197] due to the large memory footprint of the images, other normalisation layers can be adopted. For example, *layer normalisation* [198] operates along the channels (features) dimension, while *instance normalisation* [199] acts as a sample-wise *batch normalisation* (see Figure 3.7). Finally, *group normalisation* [200] operates over a group of channels for each training examples, and can be thought of as an operation in between *layer* and *instance* normalisation.

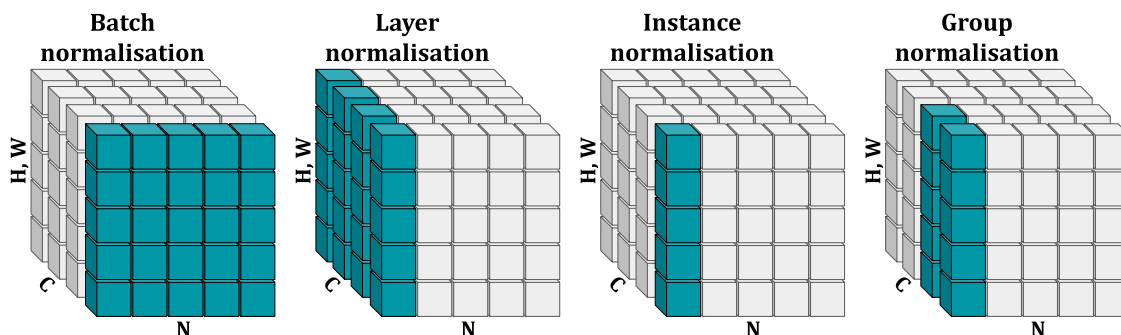


Figure 3.7: Normalisation layers where N is the number of batches, C is the number of channels and H, W are the spatial dimensions of the tensor. In each panel the teal boxes represent the pixels which will be normalised by the same mean and variance, computed from the values of these pixels. Image adapted from [200].

Optimisers. Training a neural network requires an optimiser to automatically update the model’s parameters in response to the output of the loss function. The simplest optimiser is a class *gradient descent* which updates the model’s weights θ by directly subtracting the gradient of the loss function J with regards to the weights:

$$\theta \leftarrow \theta - \eta \nabla_{\theta} J(\theta) \quad (3.6)$$

where η is called the learning rate. This can become quite slow in places where the local neighbourhood is almost flat and thus the gradients are small.

For this reason, a few variants on the classic gradient descent algorithm have been proposed in the literature. For example, in *momentum* optimisation, the gradients from the past steps are taken into account on top of the local gradient. The equation becomes:

$$\begin{aligned}\mathbf{m} &\leftarrow \beta\mathbf{m} + \eta\nabla_{\theta}J(\theta) \\ \theta &\leftarrow \theta - \mathbf{m}\end{aligned}\tag{3.7}$$

where β is introduced as a hyperparameter to prevent the momentum from growing too large.

Another popular optimiser is *RMSProp* which calculates a moving average of squared gradients to normalize the gradient itself. Thus, it manages to avoid exploding gradients by decreasing the step and vanishing gradients by increasing the step. Mathematically, it is calculated as such:

$$\begin{aligned}\mathbf{s} &\leftarrow \beta\mathbf{s} + (1 - \beta)\nabla_{\theta}J(\theta) \otimes \nabla_{\theta}J(\theta) \\ \theta &\leftarrow \theta - \eta\nabla_{\theta}J(\theta) \oslash \sqrt{\mathbf{s} + \epsilon}\end{aligned}\tag{3.8}$$

where \otimes is element-wise multiplication and \oslash is the element-wise division.

Finally, the most popular optimiser used today, *Adam* (adaptive moment estimation) [201], combines momentum with RMSProp by keeping track of the past gradients and of the past squared gradients:

$$\begin{aligned}\mathbf{m} &\leftarrow \beta_1\mathbf{m} + (1 - \beta_1)\nabla_{\theta}J(\theta) \\ \mathbf{s} &\leftarrow \beta_2\mathbf{s} + (1 - \beta_2)\nabla_{\theta}J(\theta) \otimes \nabla_{\theta}J(\theta) \\ \mathbf{m} &\leftarrow \frac{\mathbf{m}}{1 - \beta_1^t} \\ \mathbf{s} &\leftarrow \frac{\mathbf{s}}{1 - \beta_2^t} \\ \theta &\leftarrow \theta - \eta\mathbf{m} \oslash \sqrt{\mathbf{s} + \epsilon}\end{aligned}\tag{3.9}$$

where t represents the iteration number. Figure 3.8 shows a toy example of the optimisers' behaviours, for a fixed η and default hyper-parameters. Gradient descent's steps become smaller and smaller as the loss function's landscape becomes flatter, while momentum gains speed and moves downwards much faster. In fact, momentum can overshoot the minimum, which is why RMSprop and Adam are almost always preferred over the former.

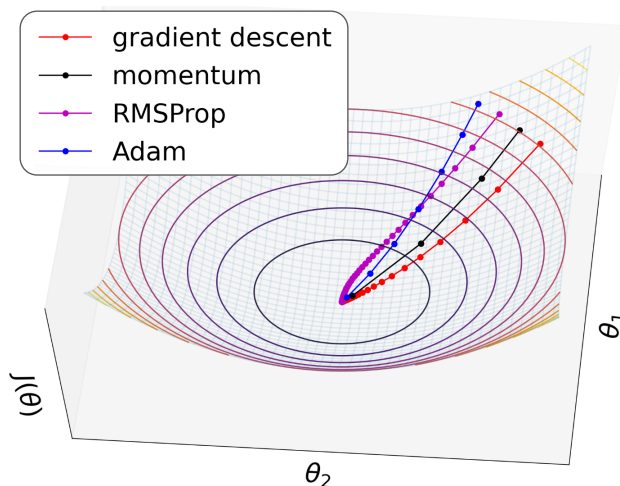


Figure 3.8: Example of four optimizers: conventional gradient descent (red), momentum (black), RMSProp (purple) and Adam (blue). Note that each example has a different starting point for visualisation purposes.

Learning rate scheduling. Choosing a good learning rate can be difficult as a too high η can make the model overshoot the minimum and diverge, while a too small η will take a long time to train. Using adaptive optimisers such as Adam or RMSProp helps, but in many cases it is preferred to have a learning rate scheduler to help the training settle down faster. For this, a few popular strategies can be used.

First, the predetermined *piecewise constant learning rate scheduling*, starts with η from a high learning rate, such as 0.1, and becomes smaller and smaller every n epochs. Second, the *exponential scheduling*, changes η as a function of the iteration number t and number of steps r . For example, $\eta(t) = \eta_0 10^{-t/r}$ will cause the learning rate to drop by a factor of 10 every r steps. Finally, the *cyclical learning rate scheduling* [202], which we also employ in our work, varies the learning rate between 2 values. Variants of this exist, where, for example, the learning rate upper threshold becomes smaller as training progresses.

Figure 3.9 shows five example learning rate schedulers. In the left figure we used log scale to show how the η changes at each iteration, starting from $\eta_0 = 0.1$. In the piecewise constant case, the learning rate was divided by 10 after every 25 iterations, while on the exponential case, the learning rate was varied with: $\eta(t) = \eta_0 10^{-t/25}$. In the cyclical learning rate case, the figure on the right shows three variants. First, the *triangular* learning rate scheduler shown in red varies η from 0.1 to 0.0001 by steadily increasing or decreasing the value every 25 iterations. Second, the *triangular2* learning rate scheduler shown in blue varies η in a similar fashion, but the learning rate difference is cut in half at the end of each cycle. Finally, the *exp_range* policy shown in magenta varies η between the two values, while decreasing the upper threshold by an exponential factor of: γ^t , where $\gamma = 0.98$ in our example figure and t is the iteration number.

Regularisation. Due to having a large number of parameters, neural networks

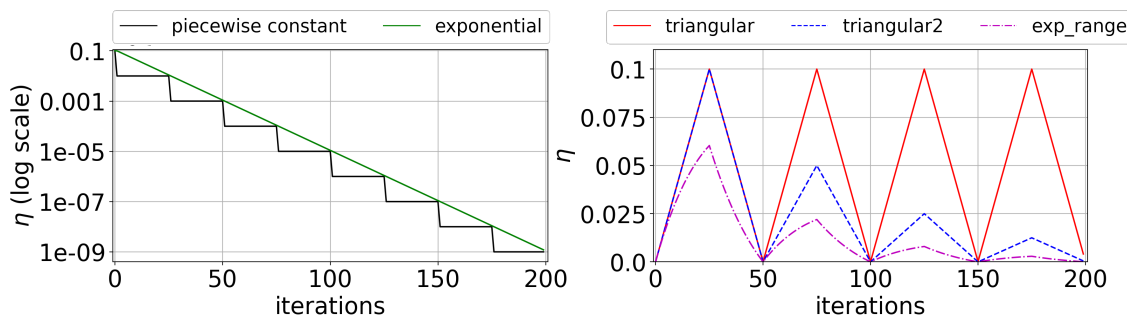


Figure 3.9: Example of learning rate schedulers: a piecewise constant learning rate scheduler (black line), exponential scheduling (green line) and three types of cyclical learning rate scheduling (red, blue dotted and purple dashed).

can exhibit overfitting behaviour [203]. To counteract this, one can feed more training data to the model, or, if unavailable, apply regularisation techniques. Some practical approaches are described in the following paragraphs.

First, *early stopping* is a technique through which both the training and the validation losses are monitored during training and the learning is stopped before the validation starts to diverge. Then, l_1 and l_2 regularisation can be applied to constrain a neural network by introducing in the loss function a penalty term (Ω) applied to the network’s weights \mathbf{W} . In the case of l_1 regularisation (also known as Lasso regularisation) the penalty is the L1 norm on the weights: $\Omega(\mathbf{W}) = \|\mathbf{W}\|_1$, while the l_2 regularisation (also known as Ridge regularisation) the penalty is the L2 norm: $\Omega(\mathbf{W}) = \|\mathbf{W}\|_2^2$. Another popular technique is *dropout*, which randomly ‘turns off’ neurons during training with some probability p . This results in a different layer node-wise and connectivity-wise every time an update occurs during training. Hinton *et al.* [204] introduced the concept of dropout in 2012 and showed improved results on a range of different applications from computer vision (*e.g.*, handwritten digit recognition).

Finally, *data augmentation* is one of the most widely used techniques today for regularisation. It consists of generating new instances by applying transformations to the images in your training set, thus boosting the available data. General methods for data augmentation include: the addition of noise, changes in image intensity (brightness, saturation and contrast), or random affine transformations (rotation, translation, scaling, shearing). In fact, the seminal paper by Ronneberger *et al.* [205] used random elastic deformations when training the proposed 3D medical image segmentation U-Net. It is worth mentioning that processing medical images poses its own difficulties which are not encountered in the computer vision world of natural images. For example, MR data is often three dimensional (or even four dimensional in higher order cases such as DTI), and includes metadata to describe the physical properties of voxels. For this reason, Pérez-García *et al.* [206] introduced TorchIO, a Python library for loading, pre-processing and medical image data augmentation. On top of the aforementioned transforms such as random affine, flip, or elastic deformations, it also includes downsampling on a particular axis (random anisotropy) [207], MRI k-space motion artifacts [208], as well as ghosting, MRI spikes, bias field

[209], image blurring and noise. It has been successfully used in many deep learning medical image tasks [210, 211] and continues to be developed and improved.

3.1.3 Network architectures

A typical network topology found in the computer vision literature is composed of convolutional layers, activation functions, max- or average-pooling layers, and, for classification tasks, fully-connected layers attached at the end to output the required predictions. For the purpose of this thesis, this section introduces three types of neural network (NN) architectures: the autoencoder (AE) and its probabilistic variant, the variational autoencoder (VAE), the U-Net and a state-of-the-art implementation called no-new-Net (nnU-Net), and the generative adversarial network (GAN) and its extension known as the Cycle-GAN. These are important components of many deep learning based applications while also representing the basic structures for the proposed models in this thesis. To date, many more neural network architectures have been created, however, it is not the purpose of this thesis to review their capabilities and for this we refer the reader to a recent survey paper on the subject [212].

AE. Autoencoders are an unsupervised learning technique which aim to efficiently learn a representation of the input data x without supervision. They can be seen as two networks: an *encoder* represented as a function $z = f(x)$, and a *decoder* represented as a function $\hat{x} = g(z)$, where z is often called the *latent space* (or *bottleneck/codings*). Intuitively, the *bottleneck* of the autoencoder is meant to ensure that the input data is not simply copied to the output, but instead impose a compressed knowledge representation of this data, thus capturing the most salient features needed to reconstruct it.

Training an AE is framed as a supervised learning problem through minimizing the error $\mathcal{L}_{\text{AE}} = \mathcal{L}_{\text{rec}}(x, \hat{x}) = \|x - g(f(x))\|^2$ between the original data and the reconstruction. Figure 3.10a shows a schematic example of an AE architecture, where both the encoder and decoder can have one or multiple hidden layers.

VAE. Variational autoencoders were introduced in 2014 by Kingma *et al.* [213]. Intuitively, VAEs can be thought of as an autoencoder whose training is regularised to ensure the latent space behaves well enough to allow generative processes. More specifically, VAEs encode the input data as a distribution over the latent space, sample from this distribution and pass it through the decoder to reconstruct it. From an architectural point of view, the encoder of a VAE produces a mean μ and a standard deviation σ . The actual coding z is then sampled randomly from a Gaussian distribution parameterised by μ and σ using the *reparametrization trick*: $z = \mu + \sigma \odot \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$.

In practice, VAEs encode the $\log \sigma^2$ instead of σ , as the logarithm function

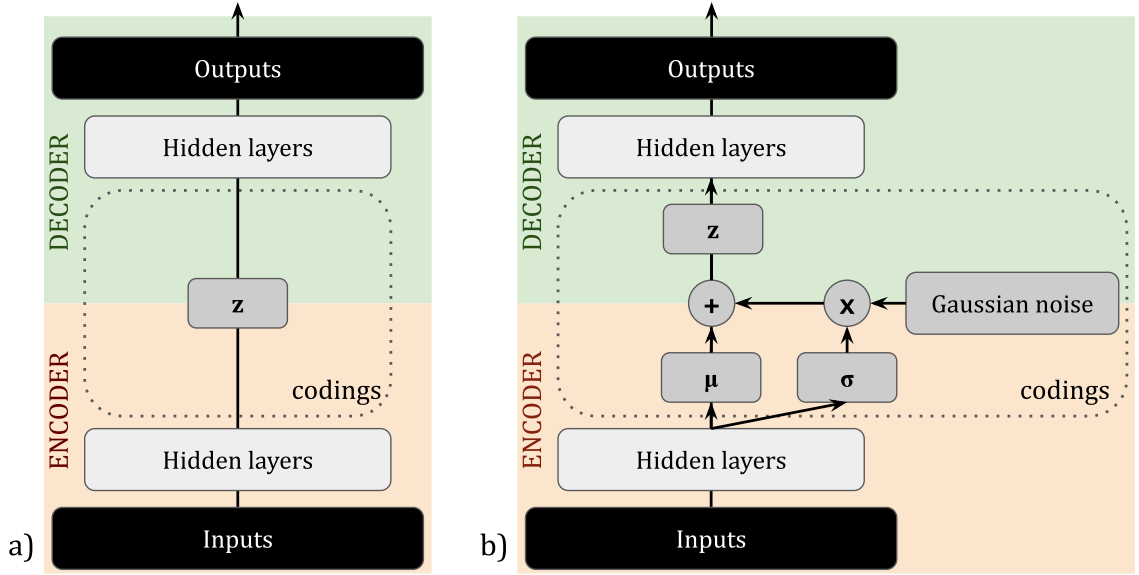


Figure 3.10: Schematic example of a) an autoencoder and b) a variational autoencoder, showing both the encoder and decoder parts of the architecture, as well as the bottleneck/codings z . Note that in the VAE case z is sampled from a Gaussian distribution parameterised by μ and σ .

has the range of set of real numbers \mathbb{R} , while the variance is constrained to be positive (i.e. $\sigma^2 \in \mathbb{R}_+$). In this case, the codings become: $z = \mu + e^{0.5 \log \sigma^2} \odot \epsilon$. VAEs are trained with a loss function which aims to minimise the reconstruction error (similar to AEs), as well as a ‘regularisation term’ (applied on the latent layer) whose objective is to impose structure on the latent space by making its underlying distribution close to a normal distribution. This is achieved through the Kullback-Leibler (KL) divergence between the distribution returned by the encoder and a standard multi-dimensional Gaussian (known as the *latent prior*). The general formula for the KL divergence for two multivariate Gaussians of dimension n is [214]:

$$\mathcal{D}_{\text{KL}}[p_1||p_2] = \frac{1}{2} \left[\log \frac{|\Sigma_2|}{|\Sigma_1|} - n + \text{tr}\{\Sigma_2^{-1}\Sigma_1\} + (\mu_2 - \mu_1)^T \Sigma_2^{-1} (\mu_2 - \mu_1) \right] \quad (3.10)$$

For the VAE case, let the encoder distribution be denoted as $q(z|x) = \mathcal{N}(z|\mu, \Sigma)$ where $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$, and the latent prior as $p(z) = \mathcal{N}(0, I)$. Then, $p_1 = q(z|x)$ and $p_2 = p(z)$, which means that $\mu_1 = \mu$, $\Sigma_1 = \Sigma$, $\mu_2 = 0$, $\Sigma_2 = I$.

Thus, equation 3.10 becomes:

$$\begin{aligned}
 \mathcal{D}_{\text{KL}}[q(z|x)||p(z)] &= \frac{1}{2} \left[\log \frac{|I|}{|\Sigma|} - n + \text{tr}\{I^{-1}\Sigma\} + (0 - \mu)^T I^{-1} (0 - \mu) \right] \\
 &= \frac{1}{2} \left[-\log |\Sigma| - n + \text{tr}\{\Sigma\} + \mu^T \mu \right] \\
 &= \frac{1}{2} \left[-\log \prod_i \sigma_i^2 - n + \sum_i \sigma_i^2 + \sum_i \mu_i^2 \right] \\
 &= \frac{1}{2} \left[-\sum_i \log \sigma_i^2 - n + \sum_i \sigma_i^2 + \sum_i \mu_i^2 \right] \\
 &= \frac{1}{2} \left[-\sum_i (\log \sigma_i^2 + 1) + \sum_i \sigma_i^2 + \sum_i \mu_i^2 \right]
 \end{aligned} \tag{3.11}$$

Training a VAE therefore consists of minimizing the sum of two losses: the generative (reconstruction) loss which compares the model prediction with the original input, and the latent loss which compares the encoder’s latent codings with a zero mean, unit variance Gaussian prior: $\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{rec}}(x, \hat{x}) + \mathcal{D}_{\text{KL}}[q(z|x)||p(z)]$. Figure 3.10b shows a schematic example of a VAE architecture.

U-Net. The U-Net is a network architecture developed in 2015 by Ronneberger *et al.* [205] and has become one of the most popular methods for deep learning (biomedical) image segmentation today. It is a type of encoder-decoder architecture which gained its name due to its symmetric shape (see Figure 3.11).

The encoder part of the network, also known as the *contracting path*, generally consists of convolutional layers (with increasing number of filters at each encoder block), followed by ReLU activation functions (blue arrows in Figure 3.11a), and a max pooling layer through which the spatial dimensions of the feature maps are generally halved (red arrows in Figure 3.11a). The decoder, also known as the *expansive path*, consists of transposed convolutions (green arrows in Figure 3.11a) whose output is concatenated, through the use of skip connections (gray arrows in Figure 3.11a), with the encoder feature maps at the corresponding level. The skip connections provide additional information to the decoder which helps it yield better features, as well as acting as a shortcut for an improved gradient flow. The last decoder step is a convolutional layer which maps the feature vectors to the desired number of classes.

Initially, the U-Net was developed for 2-D images and showcased on electron microscopy stacks of neuronal structures, as well as light microscopy images of cells [205]. In 2016, Çiçek *et al.* [215] introduced the 3-D version of U-Net, where all the convolutional and max-pooling layers were replaced with their three dimensional counterparts (see Figure 3.11b). The authors showed improved segmentation results when compared to a pure 2-D implementation. Additionally, they introduced batch normalisation between each convolutional layer and ReLU activation function (see Figure 3.11b) which further improved its performance.

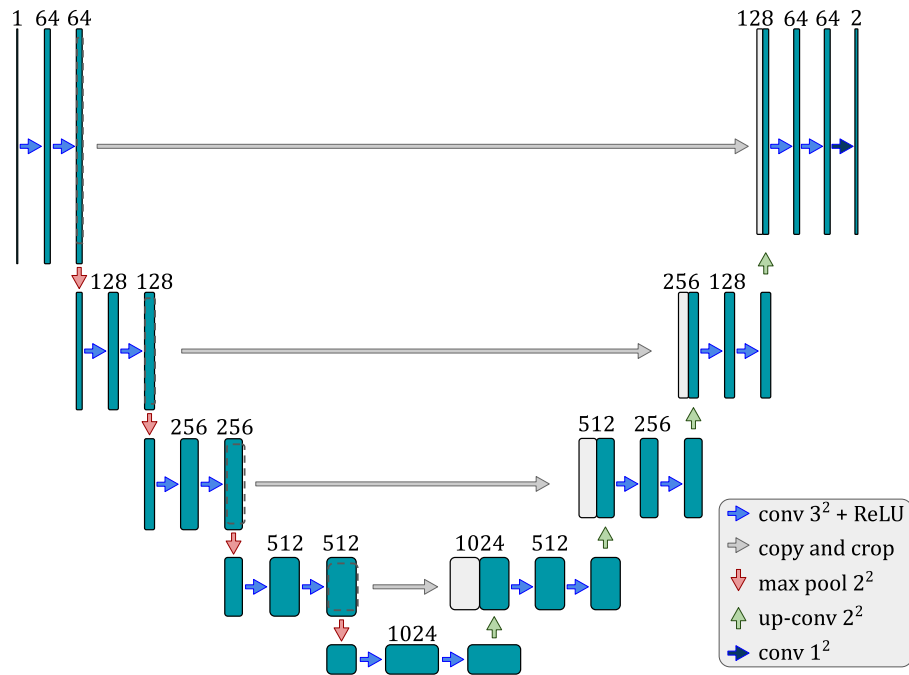
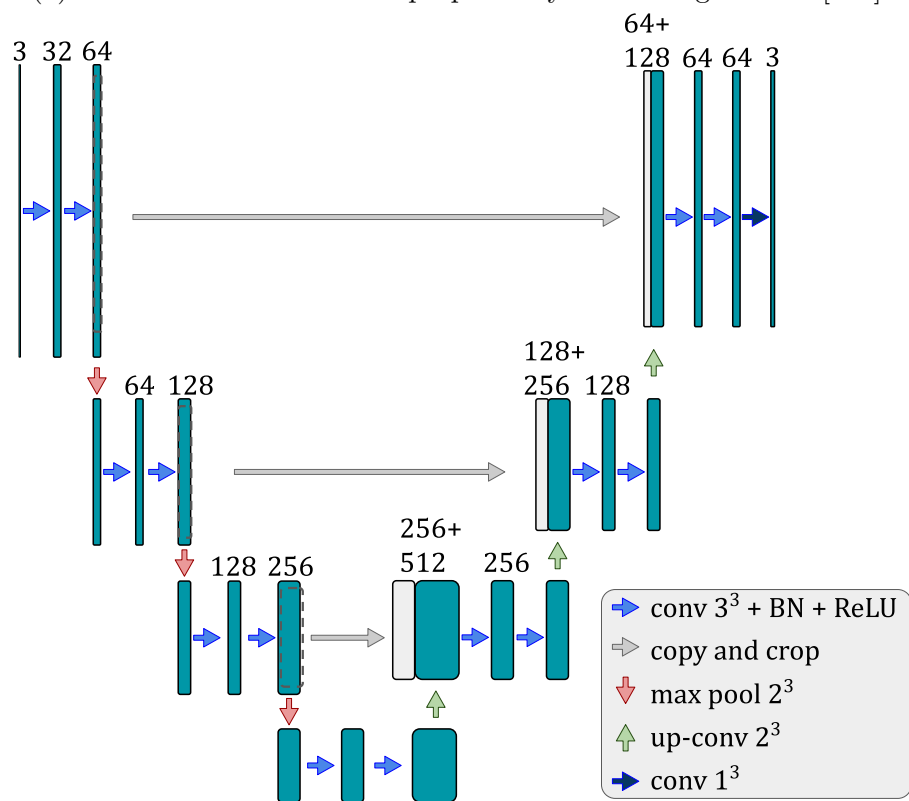
(a) 2D U-Net architecture as proposed by Ronneberger *et al.* [205].(b) 3D U-Net architecture as proposed by Çiçek *et al.* [215].

Figure 3.11: U-Net architectures for both 2D and 3D applications.

nnU-Net. A state-of-the-art implementation of the U-Net is called nnU-Net and it aims to be an out-of-the-box method for any biomedical imaging segmentation tasks. In a nutshell, the nnU-Net is a framework which automatically configures itself in terms of pre- and post-processing, as well as training parameters [216], in

three main steps.

First, the authors define a series of *fixed parameters* (see blue box in Figure 3.12) that will not change between different applications and which do not require adaptation. One example of such a design decision is the network architecture they use, and which consists of both a 2D and a 3D U-Net [205, 215], with a few changes brought to the original implementation, such as instance normalisation (instead of batch normalisation) and LeakyReLU activations. In addition to the classic U-Net, they introduce a *cascaded U-Net* architecture which is a 2-step model. In the first stage, a 3D U-Net is trained on downsampled images. Then, its predictions are upsampled to the original resolution and concatenated (as one-hot encodings) with the full resolution images. A second 3D U-Net is then trained on patches of this data with the aim of further improving the predictions.

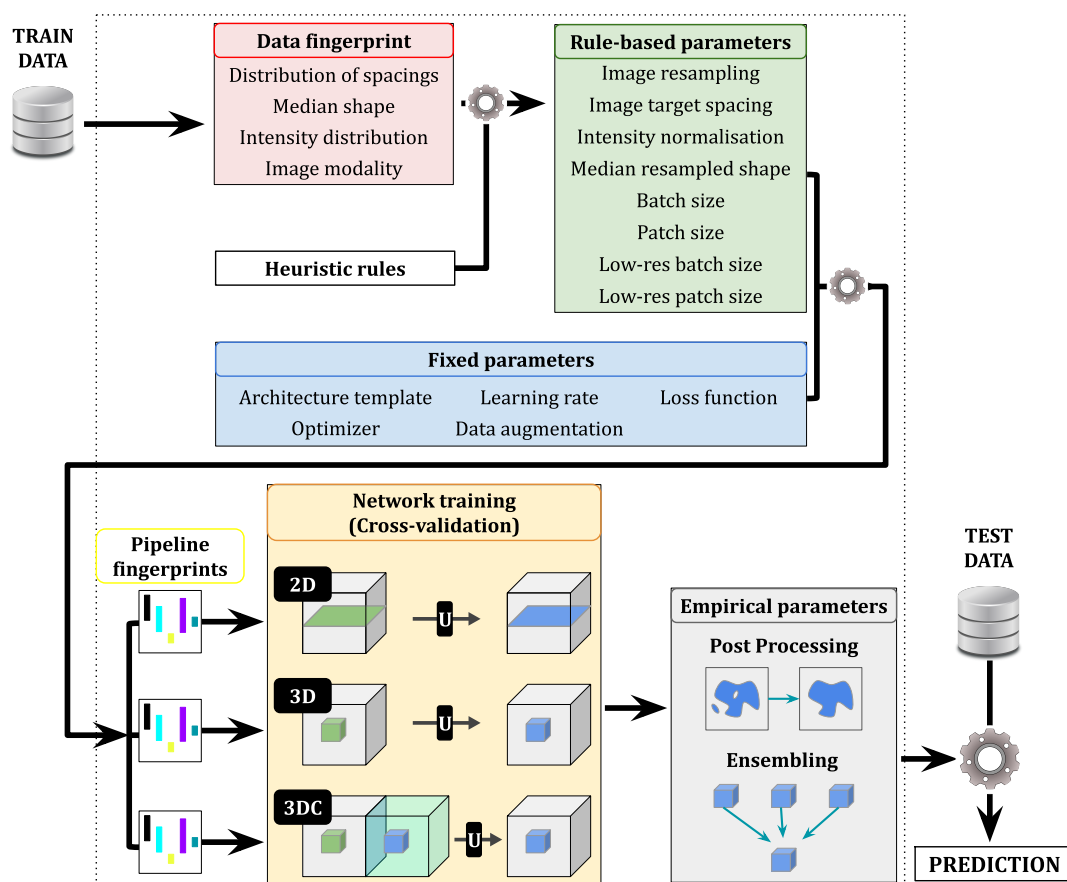


Figure 3.12: The nnU-Net pipeline showing the overall recipe for configuring a deep learning medical image segmentation solution for task-agnostic applications. While the *fixed parameters* (blue box) are not changed during training, the *data fingerprint* (red box) and their respective inferred *rule-based parameters* (green box) are dynamically changed based on the application at hand. These determine the design choices for training U-Net type architectures (both 2D, 3D and cascaded 3D U-Nets). Through cross-validation, the best performing model or ensemble of models is chosen, and, together with connected-component post-processing, deployed to make predictions on unseen test images. Image adapted from Isensee *et al.* [216].

In all cases, the loss function is kept as a combination of cross-entropy and Dice, all of the networks are trained for 1000 epochs with stochastic gradient descent, while the learning rate (initially set to 0.01) is decayed with a ‘poly’ learning rate policy [189]. Data augmentation (in the form of random affine transformations, random elastic deformations and gamma correction, among others) is also applied on the fly during training [216].

Second, the authors define a set of *heuristic rules* (more details about these can be found in both the paper and its supplementary material [216]) to operate on the training data properties and infer the training parameters. More specifically, the *data fingerprint* (red box of Figure 3.12) contains information on the type of imaging modality, the image spacing (voxel size) or the image size (number of voxels), while the *rule-based parameters* (green box of Figure 3.12) are inferred from the *data fingerprint* using the pre-defined guiding principles. These rules determine the data-dependent parameters which need adaptation for each new application, and are in place to make decisions about: the image intensity normalisation scheme (different for CT when compared to other modalities), resampling of all images into an inferred target space, adapting the patch and batch sizes to the hardware constraints (*e.g.*, the available graphics processing unit (GPU) memory). These *rule-based parameters*, together with the *fixed parameters*, generate the *pipeline fingerprints*, which are defined as all of the choices being made during method design [216].

Finally, a 2D U-Net, a 3D U-Net and a 3D cascaded U-Net are trained in a 5-fold cross-validation using the hyper-parameters determined so far. Then, the nnU-Net framework chooses which model or combination of models to use based on the foreground Dice coefficient during the cross-validation performed on the training data. One or an ensemble of two models (through averaging of softmax probabilities) is used for inference, while ‘non-largest component suppression’ is used as post-processing if needed.

The self-configuring nnU-Net framework has demonstrated leading performance on more than 20 public medical imaging datasets [216], with a broad set of modalities such as MRI, CT or electron microscopy scans, and with organs or tissues of interest ranging from brain, liver or heart, to microscopic cells. Moreover, it delivers on the promise of being an out-of-the-box tool, as it does not require manual intervention in designing task- or modality-specific configurations, and can be applied easily to new and unseen medical datasets.

3.1.4 Network training strategies

GAN. Introduced by Goodfellow *et al.* [217] in 2014, GANs have recently become popular deep learning training strategies throughout the computer vision community due to their ability to generate new data, as well as their usefulness in reducing domain shift [218]. The vanilla GAN [217] is a framework which consists of two

networks, a generator G and a discriminator D , where the generator aims to produce realistic looking images, while the discriminator tries to differentiate between real and fake (generated) samples. During training, the gradients are backpropagated from D to G , such that the generator learns to produce examples which will eventually fool D .

Figure 3.13 shows the overall architecture of the basic (vanilla) GAN [217]. The generator G takes as input a random noise vector $z \sim p_z$ (sampled from a uniform or a Gaussian distribution), and outputs a fake sample x_g . This image is expected to be similar to a real sample x_r which is drawn from the data distribution p_r . The generated samples form a distribution p_g which, through appropriate training, should be an approximation of the real data distribution. The discriminator, on the other hand, associates a probability of either being real or fake to the samples given as input. The loss functions for training the discriminator and the generator can be defined as:

$$\begin{aligned}\mathcal{L}_D^{GAN} &= \max_D \mathbb{E}_{x_r \sim p_r} [\log D(x_r)] + \mathbb{E}_{x_g \sim p_g} [\log(1 - D(x_g))] \\ \mathcal{L}_G^{GAN} &= \min_G \mathbb{E}_{z \sim p_z} [\log(1 - D(G(z)))]\end{aligned}\tag{3.12}$$

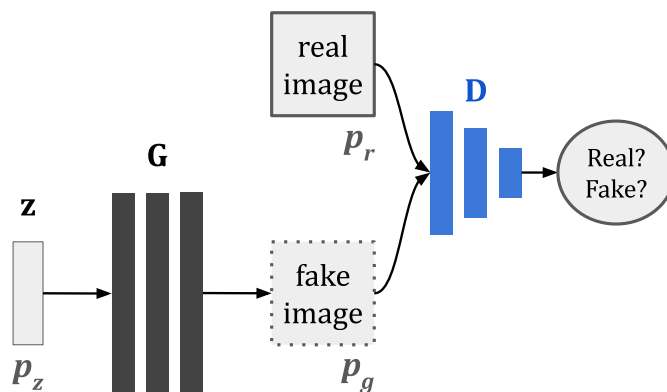


Figure 3.13: Vanilla GAN [217] where the generator G outputs a fake image ($x_g \sim p_g$) given a random vector z as input ($z \sim p_z$), while the discriminator D aims to classify the input samples ($x_r \sim p_r$, or $x_g \sim p_g$) as either real or fake.

During training, the 2 networks evolve together, while also competing against each other. One potential problem with this training objective is when one of the networks overpowers the other one. Most often, the discriminator becomes too good at distinguishing between real and generated images, thus reaching a stage where there is no gradient flow coming from it to guide training G . A second issue, which is of high importance to the medical imaging domain, is that of hallucinating features and introducing geometrical distortions in the generated images [219].

pix2pix. In the original setup, the GAN transformed noise z into sample x_g . When adding auxiliary information as input, the GAN can be extended to produce images with specific properties. In fact, Isola *et al.* [220] was the first to introduce a general purpose image-to-image translation framework called *pix2pix*. Figure 3.14 shows the overall architecture where images from domain A are translated into

domain B. This setup requires aligned paired images to ensure fidelity of generated data.

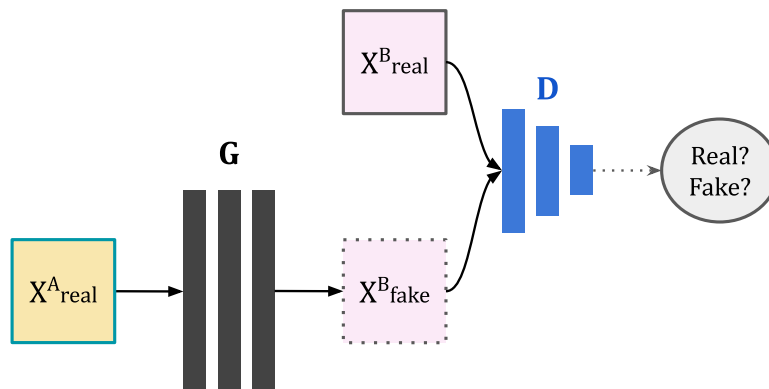


Figure 3.14: Image-to-image translation (pix2pix) [220] where the generator G outputs a fake image (X^B_{fake}) given a real image (X^A_{real}) as input. The main aim of the generator is to *translate* the real image from the source domain (A) to the target domain (B), while the discriminator D aims to classify the samples from domain B (X^B_{real} or X^B_{fake}) as either real or fake.

Cycle-GAN. To relax this constraint and allow for training with unpaired datasets, Zhu *et al.* [221] and Kim *et al.* [222] introduced the Cycle-GAN, where two generators are used to translate from one domain to another and back, while two discriminators are responsible for each domain's samples. The overall architecture is shown in Figure 3.15, where the cycle consistency loss enforces the two mappings ($A \rightarrow B$, and $B \rightarrow A$) to be reverses of each other.

In the medical imaging field, *pix2pix*-type frameworks are generally used when paired data is available, while Cycle-GAN-type models are used for more general applications. This is because the latter can more easily constrain the generated data to be anatomically correct, and, in fact, Wolterink *et al.* [223] found that training a Cycle-GAN model with unpaired images was better than with paired data at synthesising MR to CT images. On the other hand, Zhang *et al.* [224] found that cycle consistency was not sufficient to ensure the lack of geometric distortions, and so they introduced segmentation networks to provide shape constraints to the translated anatomies.

In *pix2pix*-type frameworks, constraints have been added through regularisation terms in order to preserve anatomy between fake and real images, while also using unpaired datasets. For example, Mahmood *et al.* [225] introduced a self-regularisation loss when generating synthetic representations of real endoscopy images. This term was added to the overall generator loss and was defined as the l_1 norm between the original and the synthesized images. Finally, BenTaieb *et al.* [226] introduced an edge-weighted regularisation term to help the generator preserve edge features between the input and the synthesized images.

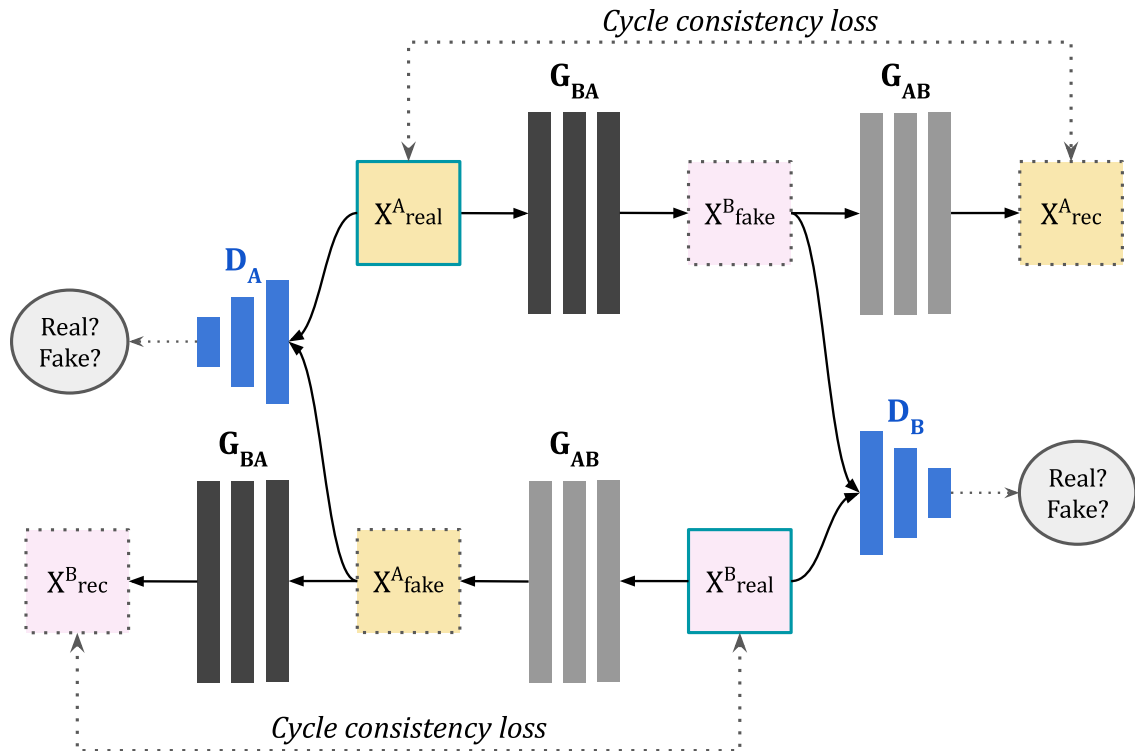


Figure 3.15: Cycle-GAN [221] where two generators G_{AB} and G_{BA} are used to translate images from domain A to domain B, and from domain B to domain A, respectively. For this, 2 discriminators (D_A and D_B) are used as classifiers for each of the 2 domains.

3.2 Deep learning for medical image analysis

3.2.1 Deep learning-based medical image registration

This section focuses on reviewing the **deep learning based registration algorithms** literature. The methods covered here are grouped into three categories, a taxonomy based on the survey paper by Haskins *et al.* [227].

Deep iterative registration algorithms

The class of deep learning algorithms that fall under the *deep iterative registration* umbrella use networks as a means to augment the performance of an iterative, intensity based classic registration framework. This is shown schematically in Figure 3.16, where solid lines represent information flow during training and inference, while dashed lines are for training only. These methods, labelled as **deep similarity based registration** by Haskins *et al.* [227], use deep learning to estimate a similarity metric, which is then introduced in a classic registration framework. In the remainder of this section, some notable papers from the field are presented.

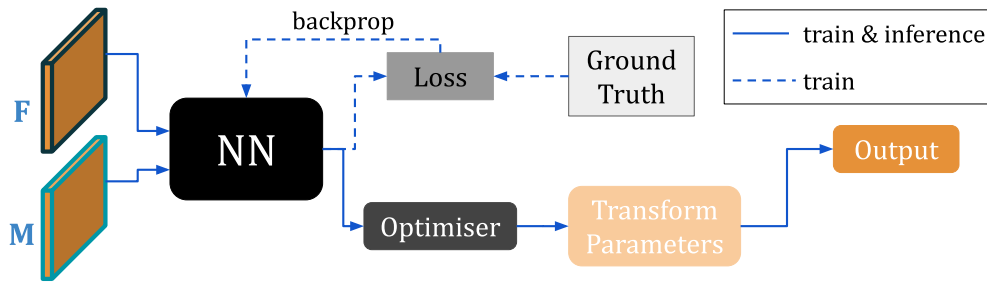


Figure 3.16: Overall schematic representation of **deep similarity based registration** methods where a neural network (shown in the figure as the black-box **NN**) is used to estimate a similarity metric, which is then introduced in a classic registration framework. Note that both the optimisation algorithm (*Optimiser*), and the transformation model (*Transform Parameters*) are part of a classical intensity-based registration framework. Image adapted from Haskins *et al.* [227].

First, Wu *et al.* [228] constructed a convolutional autoencoder (CAE) network to extract data-adaptive features from 3D MRI image patches. Pairs of hierarchical features are then used in a classical registration framework where gradient descent optimizes the NCC between them. Their method outperformed HAMMER [229] and Demons [97]. Eppenhof *et al.* [230] used a CNN to estimate the registration error between pairs of 3D thoracic CT scans. Their network was trained on synthetically deformed image patches and evaluated on deformable registrations of inhale-exhale pairs of thoracic CT scans [230]. Similarly, Simonovsky *et al.* [231] constructed a network to estimate the image similarity for multimodal registration. More specifically, they used a CNN to compute the dissimilarity between 3D T_1 w and T_2 w brain MRI volumes. Using this, a classical registration framework was able to find the parameters of a deformation field that registered the two modalities better than using MI as a similarity metric. Finally, Wright *et al.* [232] used a long short-term memory (LSTM) spatial transformer network to register MRI and ultrasound (US) volumes. Their method dealt with global affine registrations and it outperformed a previous multimodal image-registration algorithm [233].

Supervised transformation estimation

As the previously described methods were slow, deep learning registration research started to focus its attention towards developing faster methods. For this, supervised and semi-supervised networks were created where ground truth deformation fields or tissue segmentations were used to drive the learning process.

Initially, networks were trained using a **full supervision** approach. Figure 3.17 shows the general architecture of these methods, where the neural network predicts the transformation parameters needed to align the images. As the name suggests, ground truth deformation fields are needed during training.

Yang *et al.* [234] created a U-Net like architecture to predict the initial momenta

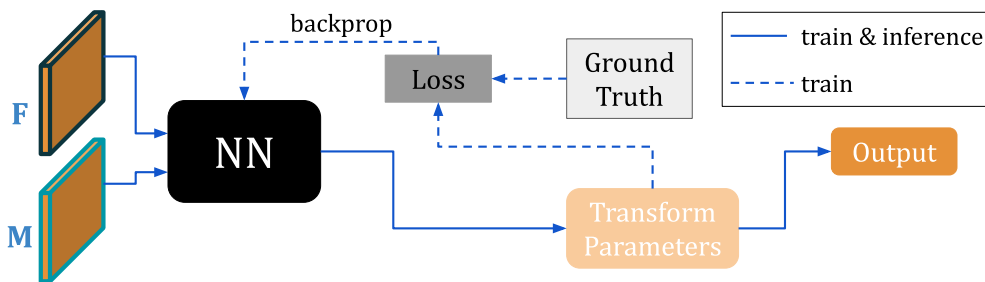


Figure 3.17: Overall schematic representation of **fully supervised deep learning registration methods** where ground truth transformation parameters are needed to drive the learning process. Image adapted from Haskins *et al.* [227].

needed to generate a deformation field through LDDMM shooting. Their network was trained on pairs of 2D or 3D MRI patches and ground truth initial momenta were computed through numerical optimization of the LDDMM shooting formulation. This method sped up computational time to only a fraction of the time required by classic optimisation-based techniques. Similarly, Rohe *et al.* [235] used pairs of 3D cardiac MR images to predict a SVF. Ground truth data was generated by computing deformation fields from mesh segmentations of image pairs.

Cao *et al.* [236] used a CNN to regress a displacement vector for given input 3D image patches. The patch-wise displacements were then aggregated into a dense deformation field by thin-plate spline (TPS) interpolation. Some of their contributions also included the sampling strategy of the patches (equalized active-points guided sampling strategy) which ensured that patches with higher gradient magnitudes and displacement values were sampled more frequently. The ground truth deformation fields used to train the network were generated by first registering the images using Syn [60] and diffeomorphic Demons [97]. Finally, Uzunova *et al.* [237] used a FlowNet [238] architecture on 2D brain and cardiac MR images. Unlike the previous methods, the ground truth deformations were estimated using statistical appearance models (SAM).

Dual supervision models train the networks on both ground truth deformation fields and a similarity measure (*i.e.*, SSD, NCC, NMI) which compares the fixed and the warped moving images. For example, Fan *et al.* [239] used a hierarchical dual-supervised fully convolutional neural network (FCNN) to predict the deformation field needed to register pairs of 3D MR images. For training, the authors used both the similarity between predicted and ground truth deformations, as well as the similarity between the moved and fixed images.

Finally, **weakly supervised** deep learning methods use tissue segmentations to drive the registration. A notable example is by Hu *et al.* [183, 240] where label similarity is employed to train a network on pairs of MRI and transrectal ultrasound (TRUS) data. In this work, the authors develop two networks: a *Global-net* for estimating an affine transformation and a *Local-net* for predicting the dense

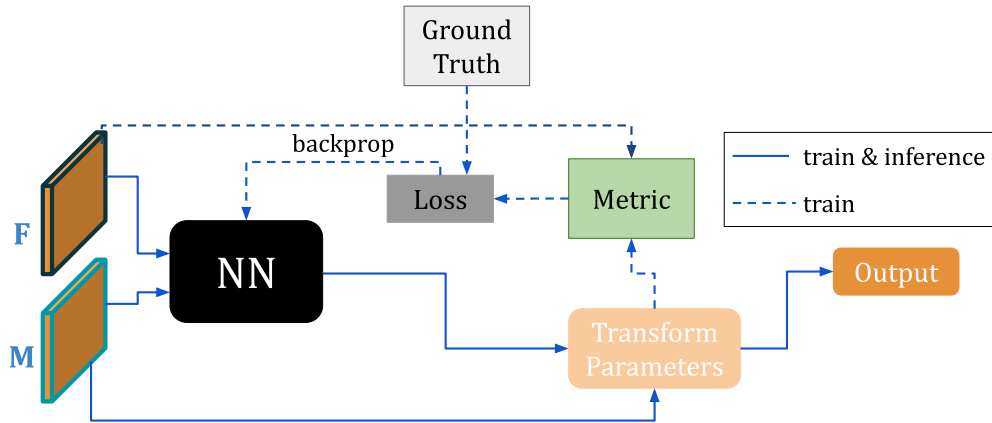


Figure 3.18: Overall schematic representation of **dual supervised deep learning registration** methods where, on top of using ground truth deformation fields to drive the learning process, an image similarity metric is employed to measure the error between the warped moving and the fixed images. Image adapted from Haskins *et al.* [227].

displacement field. The inputs to the combined (*Global/Local-net*) network are pairs of MR-TRUS 3D volumes and their respective organ segmentation maps, while training is done by minimizing the Dice loss between the moved and fixed labels. Figure 3.19 shows the proposed architecture during training and inference, which highlights one of the novelties of this work: labels are not needed at test time, only during training.

Later, Hu *et al.* [241] developed an adversarial network to perform the MR-TRUS registration while simultaneously maximizing label similarity and minimizing an adversarial loss on the deformation field. This regularisation strategy outperformed standard bending energy based regularisation. Hering *et al.* [242] proposed a U-Net like architecture to register pairs of 2D cine-MR images. In their work, the authors introduced a loss function that took into account both the similarity between the fixed and moved MR images (as edge-based normalized gradient fields distance measure) and their respective labels (as SSD).

Unsupervised transformation estimation

Although the previous methods performed well and provided high quality registrations, the need for ground truth data meant that pre-processing was needed for training the networks. For this reason, research moved towards unsupervised methods, which bypassed the need for collecting or simulating data.

Unlike dual supervised models (see Figure 3.18), **similarity metric based unsupervised methods** do not employ ground truth transformation parameters, and train their models on a well-defined similarity measure only (see Figure 3.20). Li *et al.* [243, 244] were one of the first to create and train a CNN model for deformable

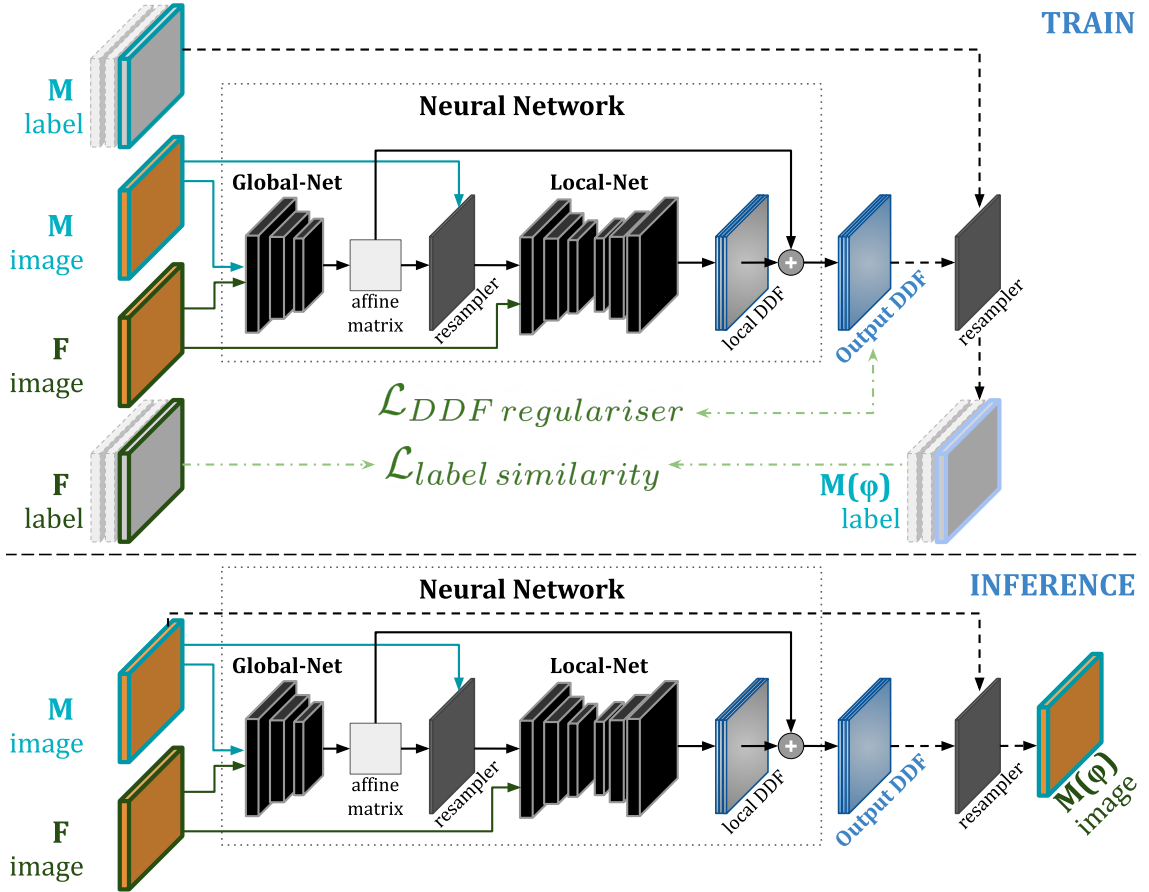


Figure 3.19: Weakly supervised deep learning for multi-modal deformable registration during training and inference. The inputs to the network are pairs of T_2W MR and TRUS images, while label data is used to compute the loss between the warped and the fixed labels. Bending energy [68] is added to the overall loss to regularise the predicted output dense displacement field. At inference time, the labels are no longer needed, as the network predicts the dense displacement field using pairs of image data only. Figure adapted from Hu *et al.* [183]. For simplicity, the Global-/Local-Net architectures are not explicitly shown in the figure, however, the reader can consult [240, 183] for their detailed description.

registration of 3D brain MRI volumes. Their loss function was composed of an NCC similarity measure between the fixed and the warped images, and a smoothing constraint on the predicted deformation field. Their work outperformed ANTs [245]. Similarly, de Vos *et al.* [246] trained a CNN using an NCC similarity metric on 4D cardiac cine MR volumes. The following year, they improved their method by constructing a multi-stage, multi-scale registration network [247] to perform an initial affine and a subsequent B-spline non-linear registration between images of the same modality. Their method outperformed a classic registration technique called Elastix [248]. Stergios *et al.* [249] used a CNN to jointly predict a linear (affine) and a non-linear transformation capable of registering inhale-exhale pairs of lung MR volumes. The loss function was made up of a MSE metric and regularization terms, and it outperformed ANTs-based [60] registration.

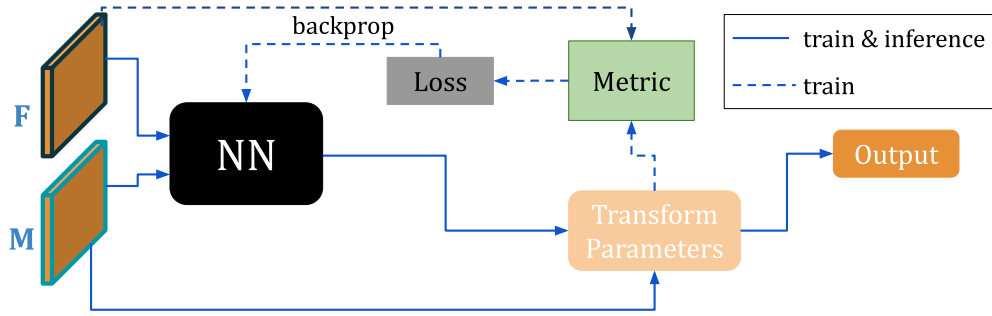


Figure 3.20: Overall schematic representation of **unsupervised deep learning registration** methods where no ground truth data is used and an image similarity metric is employed to measure the error between the warped moving and the fixed images. Image adapted from Haskins *et al.* [227].

Balakrishnan *et al.* [250, 182] introduced Voxelmorph as a general unsupervised image registration framework. In their work, they employed a U-Net like architecture trained to generate a dense deformation field (see Figure 3.21). The loss function was made up of an image similarity metric (LNCC) and a regularisation term. Following

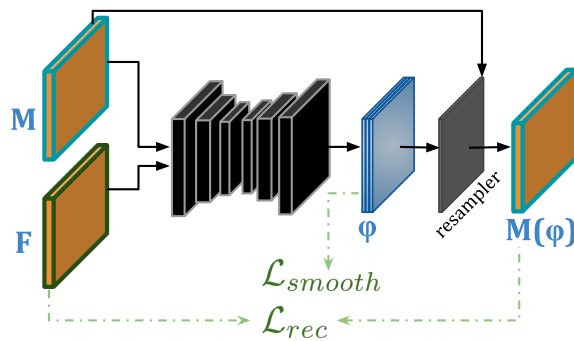


Figure 3.21: Voxelmorph unsupervised image registration framework as proposed by Balakrishnan *et al.* [250, 182]. A pair of moving (\mathbf{M}) and fixed (\mathbf{F}) images form the input to a U-Net-like architecture trained to generate a dense displacement field φ . A spatial transformer layer resamples the moving image into the warped image $\mathbf{M}(\varphi)$. The loss function is made up of an image similarity metric between the fixed and warped images, and a diffusion regularizer on the predicted displacement field to encourage smooth deformations. For simplicity, the U-Net architecture is not explicitly shown in the figure, however, the reader can consult [250, 182] for a detailed description.

this work, Dalca *et al.* [181, 251] extended Voxelmorph to predict the deformation field through variational inference (see Figure 3.22). Moreover, it introduced *scaling and squaring* layers to ensure a diffeomorphic deformation. Their work outperformed ANTs-based registration [60].

Krebs *et al.* [252] had a different approach to a variational framework, and introduced a conditional variational autoencoder (CVAE) network to learn a probabilistic model for image registration. Similar to Dalca *et al.* [181], their proposed network outputs a velocity field v which is then integrated to produce a deforma-

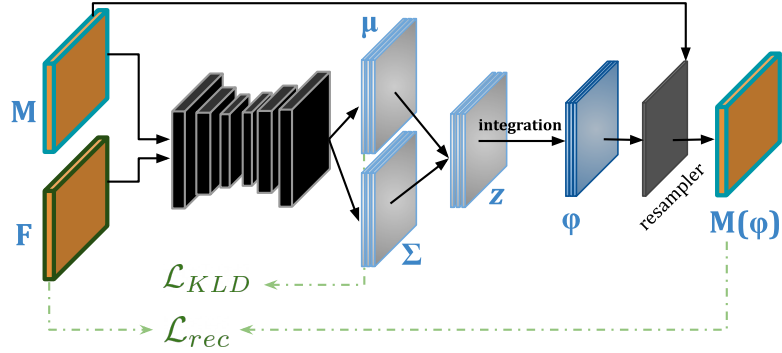


Figure 3.22: Diffeomorphic Voxelmorph unsupervised image registration framework as proposed by Dalca *et al.* [181]. The network outputs the approximate posterior probability parameters, *i.e.*, velocity field mean μ and variance Σ , from which the velocity field \mathbf{z} is sampled. Through *scaling and squaring* layers \mathbf{z} is then transformed into a topology-preserving deformation field φ . The loss function is made up of an image similarity term (\mathcal{L}_{rec}) which encourages the warped moving $M(\varphi)$ to be similar to the fixed image F , and the KL divergence (\mathcal{L}_{KLD}) which encourages the posterior to be close to a multivariate normal prior. For simplicity, the U-Net architecture is not explicitly shown in the figure, however, the reader can consult [181] for a detailed description.

tion field φ . In their work, the moving image acts as the conditioning data and is warped to match the fixed image. The loss function is therefore comprised of the reconstruction loss between the warped moving and the fixed images (as LNCC) and the KL divergence between the encoded latent distribution and a prior probability distribution which acts as a regularisation term. Later, the authors introduced a multi-scale approach [253] where estimations of velocity fields, deformation fields and deformed moving images were generated at three different scales: the original size (*full scale*), half of the original size (*middle scale*) and a quarter of the original size (*coarse scale*). Their experiments showed that the multi-scale approach led to improved registration results when compared to their previous network.

A different multi-scale approach was brought forward by Kuang *et al.* [254]. The authors introduced inception modules (see Section 3.2.2 for a description of inception modules) into their CNN architecture for the purpose of capturing information at different spatial scales. Their network was trained using the NCC similarity metric and a regularization term, and it outperformed Voxelmorph [182]. Later, Zhang *et al.* [255] introduced inverse-consistency in their proposed CNN network. More specifically, they generate both the deformation from the moving image to the fixed image and the deformation from the fixed image to the moving image, and then employ a constraint which ensures that the predicted flows are consistent. Their work outperformed ANTs-based registration [60]. Fan *et al.* [256] introduced a GAN-based approach to assess the similarity between the fixed and the moved images. Unlike other approaches where a similarity metric is explicitly introduced for the application at hand, in this work the discriminator is trained to assess the quality of the alignment. This approach outperformed diffeomorphic Demons [78] and ANTs-based [60] registration approaches.

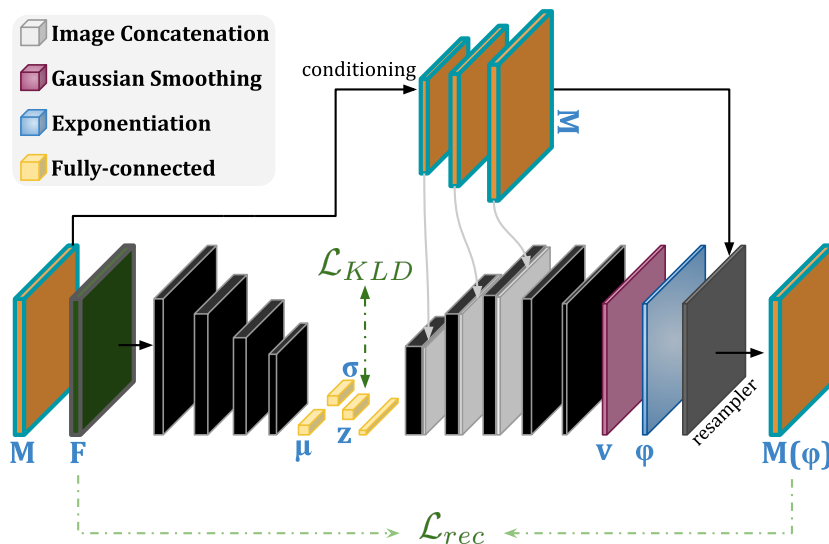


Figure 3.23: Diffeomorphic unsupervised probabilistic registration network framework as proposed by Krebs *et al.* [252]. The network outputs a velocity field v which is smoothed through a Gaussian smoothing layer, and then transformed into a topology-preserving deformation field φ through the exponentiation (*scaling and squaring*) layer. The decoder part of the network is conditioned on the moving image \mathbf{M} at three different scales (the original size and 2 downsampled versions). The loss function is made up of an image similarity term (\mathcal{L}_{rec}) which encourages the warped moving $\mathbf{M}(\varphi)$ to be similar to the fixed image \mathbf{F} , and the KL divergence (\mathcal{L}_{KLD}) which encourages the vector \mathbf{z} to follow a prior distribution $p(z)$, defined as a multivariate unit Gaussian $p(z) = \mathcal{N}(0; I)$ (where I is the identity matrix). For simplicity, the CVAE architecture is not explicitly shown in the figure, however, the reader can consult [252] for a detailed description.

In 2020, Mok *et al.* [257] introduced the Laplacian pyramid image registration network (LapIRN), which utilizes a multi-resolution strategy for large deformation image registration tasks. The authors trained their framework through a coarse-to-fine scheme, by first training the network at the coarsest level alone, and then adding the next levels into training until the finest resolution is reached. Their proposed model won the 2020 MICCAI Learn2Reg¹ challenge.

So far, the focus was on pairwise registrations only. For groupwise registration, van der Ouderaa *et al.* [258] introduced GroupMorph, an extension to Voxelmorph [181], to register multiple images simultaneously. The authors stacked a series of input images along the channel axis, and trained the network to generate multiple velocity fields. Thereafter, the velocity fields are transformed into deformation fields through *scaling and squaring* layers. Finally, the authors train the network in two ways: *all-to-one*, by considering one of the inputs as the fixed and the rest of the inputs as moving, and *all-moving* by registering all of the input images to their geodesic average. The experiments they conduct show that their proposed strategy was able to simultaneously register multiple scans while preserving the performance of the original Voxelmorph [181]. Similarly, Gu *et al.* [259] proposed the symmetric

¹learn2reg.grand-challenge.org/Learn2Reg2020

cycle consistency network, where both inverse- and cycle-consistency is introduced. The inverse-consistency ensures pair-wise registrations are bidirectional, while the cycle-consistency is an extension for multiple images. More specifically, the group-wise consistency is a penalty introduced in the loss function for groups of more than three images. It ensures that for n ordered images, the composition of all the pair-wise deformations ($\varphi_{1\rightarrow 2}, \dots, \varphi_{n-1\rightarrow n}$) is equal to directly registering the first and the last images.

One common denominator to the unsupervised deep learning medical image registration methods presented so far is the use of image-based similarity metrics. A different approach, known as the **feature based unsupervised transformation estimation** method, aims to use learnt features to align images. For example, Yoo *et al.* [260] train an AE to reconstruct electron microscopy images. Then, the authors employ a spatial transformer network to align pairs of images, by minimizing the L_2 norm between the encoded features of the fixed and the warped images. In this way, the authors do not use similarity metrics in image space, but align the images based on their encoded representations.

Similarly, Lee *et al.* [261] introduce the image-and-spatial transformer network where the moving and the fixed images are inputs to an image transformer neural network which aims to produce image representations optimised for downstream tasks. These intermediate representations are then fed into the spatial transformer network. The authors also use segmentation maps to guide the registration, and show that their proposed method outperforms the unsupervised (no labels) and supervised (with labels) transformer-only network.

3.2.2 Deep learning-based medical image segmentation

This section focuses on reviewing state-of-the-art **deep learning segmentation** models, using a taxonomy proposed by Wang *et al.* [262] which groups different works based on: the backbone network architecture, the selection of network blocks, improvements brought to the training loss function, as well as the use of data augmentation.

Network architectures

One way of grouping deep learning image segmentation models is by the improvements they bring on the network architectures. This section presents a review of some of the most popular architectures found in the state-of-the-art medical image analysis literature.

U-Net. The encoder-decoder structure of the U-Net [205, 215] (see also Sec-

tion 3.1.3 U-Net) is one of the most widely used network architectures today. In fact, the U-Net is often regarded as a benchmark for many medical imaging segmentation tasks, as well as an important starting point for many network architectures [216, 263, 264, 265, 266, 267, 268].

A notable improvement to medical image segmentation applications is the **cascaded** model. In a nutshell, this strategy aims to increase segmentation accuracy by training two or more networks. In fact, the nnU-Net presented in Section 3.1.3 sometimes makes use of this scheme to further refine its predictions. In their case, a 2-step process is employed: first, a 3D U-Net is trained on downsampled biomedical images to predict segmentation maps, then, a second network (also a 3D U-Net) is trained on patches of the full resolution images concatenated with the first stage predictions with the aim of improving the predictions. Christ *et al.* [263] has a similar approach where they propose a liver and tumour segmentation cascading model as seen in Figure 3.24. One U-Net is trained to localise an organ of interest (the liver in their case), while a second U-Net is trained on images masked with the stage 1 predictions to segment smaller structures (tumours).

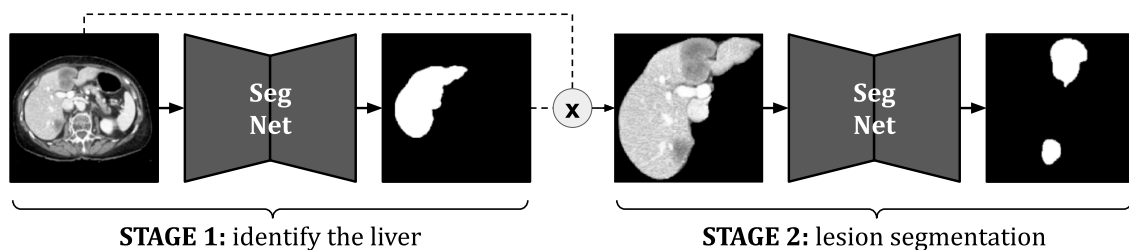


Figure 3.24: An example of an image segmentation cascading model as proposed by Christ *et al.* [263], where in the first stage a U-Net is used to segment the liver in a CT slice, while in the second stage the network takes the cropped and masked liver image as input with the aim of segmenting tumour lesions. Image adapted from [263].

HighResNet. A different approach to the U-Net was studied by Li *et al.* [269] where the goal was to design a high-resolution network capable of segmenting small structures in 3D medical images. HighResNet uses dilated convolutional layers and residual connections, and keeps the original resolution of the input volume throughout the network (see Figure 3.25).

The authors apply their proposed solution for segmenting 3D T_1w MR brain images of healthy volunteers from the ADNI² dataset into 155 structures and 5 non-brain tissues. They show improved results over U-Net and other state-of-the-art 3D biomedical image segmentation networks [197, 270], while using fewer network parameters. However, the drawback of this architecture is that, because it does not downsample the input data at any of its layers, its GPU memory footprint can become large.

DeepMedic. Driven by the need for better and more accurate medical image

²<https://adni.loni.usc.edu/>

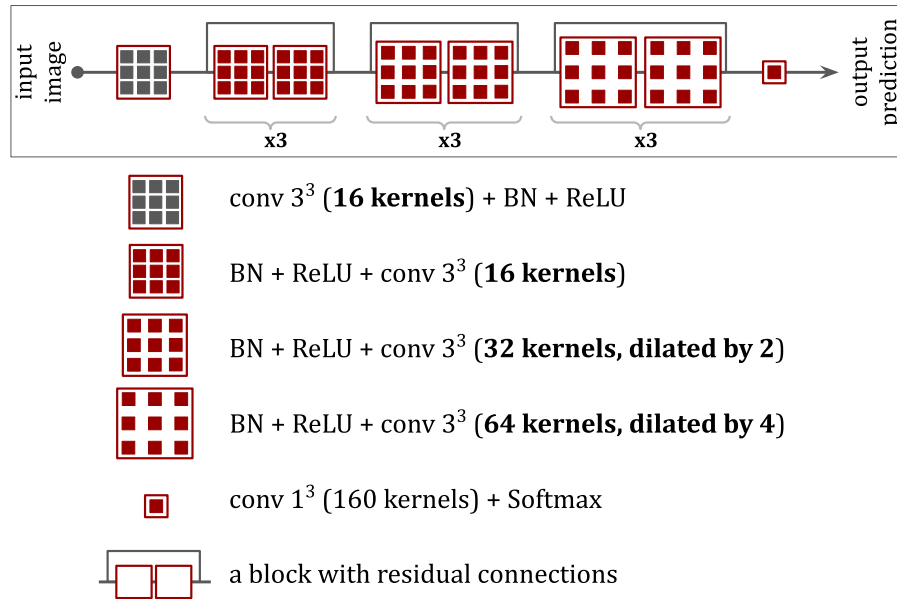


Figure 3.25: Architecture of HighResNet [269]. The first convolutional layer uses 16 filters of kernel size 3^3 , followed by batch normalisation and ReLU activation functions. Then, blocks with residual connections between their inputs and their respective outputs made up of batch normalisation, ReLU activations and (dilated) convolutions of increasing number of filters are repeated 3 times each. Finally, a 1^3 convolution followed by a Softmax activation outputs the 160-channels prediction. Image adapted from [269].

segmentation models, different approaches to improving the network architecture have been developed [271]. For example, Kamnitsas *et al.* [270] introduced a dual pathway 3D CNN architecture named DeepMedic. The main aim of this approach was to incorporate both normal and lower resolution patches in order to increase the contextual information of the model.

Their work was intended for segmenting abnormalities, application in which context is generally considered important [271]. As feeding large patches to a network can become memory expensive, their proposed solution was therefore to add a down-scaled representation of a larger context around the normal resolution patch. This is shown in Figure 3.26 where the independent streams are based on the two input patches. These are fed through different convolutional layers, while the low resolution stream is upsampled to match the normal resolution feature maps prior to becoming input to the fully connected 1^3 convolutional layers.

Dolz *et al.* [272] introduced a 3D fully convolutional network with the aim of predicting segmentation maps of subcortical brain structures in 3D MRI. Similar to DeepMedic [270], the authors modelled both local and global context, but increased the depth of the model by using smaller kernel sizes, and removed the dual pathway in favor of injecting intermediate-layer feature maps in the final prediction. They evaluated their method on the ABIDE³ dataset [273] and showed its robustness to

³The Autism Brain Imaging Data Exchange

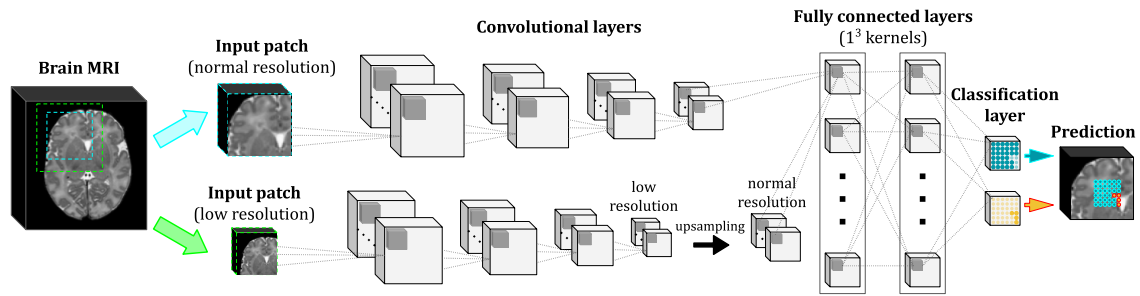


Figure 3.26: The overall architecture of the DeepMedic image segmentation model as proposed by Kamnitsas *et al.* [270], where both local (normal resolution) and larger (lower resolution) contextual information are processed simultaneously through the two convolutional pathways. For the sake of simplicity, the number of feature maps and their size has been omitted in the figure, but the reader can refer to the original paper for more details [270]. Image adapted from [270].

different ages, acquisition sites and diagnosis groups.

Network blocks

Another approach to biomedical image segmentation tasks using deep learning methods is to introduce novel network blocks. In this subsection the focus is on three types of modules which have brought improvement to segmentation networks.

Residual connections. First, residual connections have become very popular since their introduction in the computer vision field through the ResNet architecture [274] where He *et al.* showed that they are a simple, yet very effective way of easing training of deep neural networks. The problem that was explored by the authors was that of increasing the depth of a neural network, which showed that training accuracy becomes saturated and starts to degrade [274]. Introducing residual connections helped overcome this problem and the rationale for this has been explored by Veit *et al.* [275] where they showed that such networks behave like an ensemble of shallower models.

In a nutshell, a residual connection provides a way for data to skip layers and reach deeper parts of the neural network. Figure 3.27 schematically shows an example of how this can be achieved. In fact, Milletari *et al.* [197] proposed the V-Net as a 3D medical image segmentation network which incorporates residual connections, and applied it to predict segmentation maps of the prostate in MRI volumes. Residual connections were also introduced in Voxresnet for 3D MR image brain segmentation [276], 2D Res-UNet for retina vessel segmentation [277], or SEGANet where, on top of the residual units, the authors introduced instance layer-normalization and PReLU activation functions in a standard 3D U-Net and applied to segment the left atrium in short-axis dynamic cardiac MRI volumes [278].

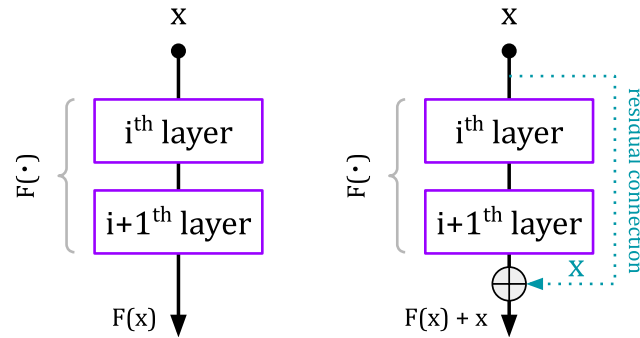


Figure 3.27: Residual connection as proposed by He *et al.* [274] is a simple and effective way of easing training of deep NNs. On the left, two layers are represented through the mapping $F(\cdot)$ which takes as input x and produces $F(x)$ as its output. On the right, the residual connection is portrayed as a bypass of these layers such that the output becomes $F(x) + x$. Image adapted from [274].

Dense connections. A similar approach to residual units was introduced by Huang *et al.* [279], where the input from each layer comes from the output of all previous layers. As each layer has access to the loss function’s gradients, networks which employ dense connections have enhanced information flow which makes them more compact while achieving better performance [279, 280]. A schematic representation of a 4-layer dense connection is shown in Figure 3.28, where the output of the l^{th} layer is concatenated to the input of the $(l + 1)^{\text{th}}$ layer. Consequently, each layer receives as input the feature maps of all previous layers.

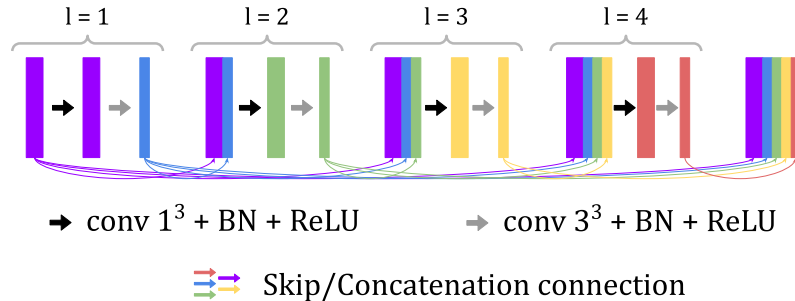


Figure 3.28: A 4-layer dense connection example showing how each layer’s output is concatenated to the next layer’s input. The input of the first block is also an input to each subsequent layers. The output is the aggregation of all previous layers’ outputs together with the original input. Image adapted from [280].

In fact, Guan *et al.* [280] proposed an improved 2D U-Net which used dense connections at each block. Moreover, Zhou *et al.* introduce U-Net++ [264, 281] where the encoder and the decoder layers are linked through a series of dense connections. They apply their proposed method on 2D and 3D datasets with the aim of segmenting the liver and lung nodules, respectively. Additionally, they test their model on 2D microscopy images and RGB videos of colon polyps. In all cases they achieve better results than the simple U-Net. Dolz *et al.* [282] introduced HyperDenseNet, a 3D multi-modal fully convolutional neural network that uses dense connections, and showed state-of-the-art results on segmenting isointense T_1w and T_2w brain MR images of 6-months old infants [283]. This model was later applied on the dHCP dataset and showed improved results over other segmentation models [284]. It is

worth mentioning, however, that although dense connections can be helpful, they can also increase the number of parameters of a network which would require more computational power.

Inception. The inception module was introduced by Szegedy *et al.* [285] as a means of allowing multiple filter sizes to co-exist in a single block. The outputs of each layer of different sized filters is then concatenated and passed to the next layer in the network. Figure 3.29 shows this schematically in its so-called ‘naive’ form (a), as well as a more complex version (b).

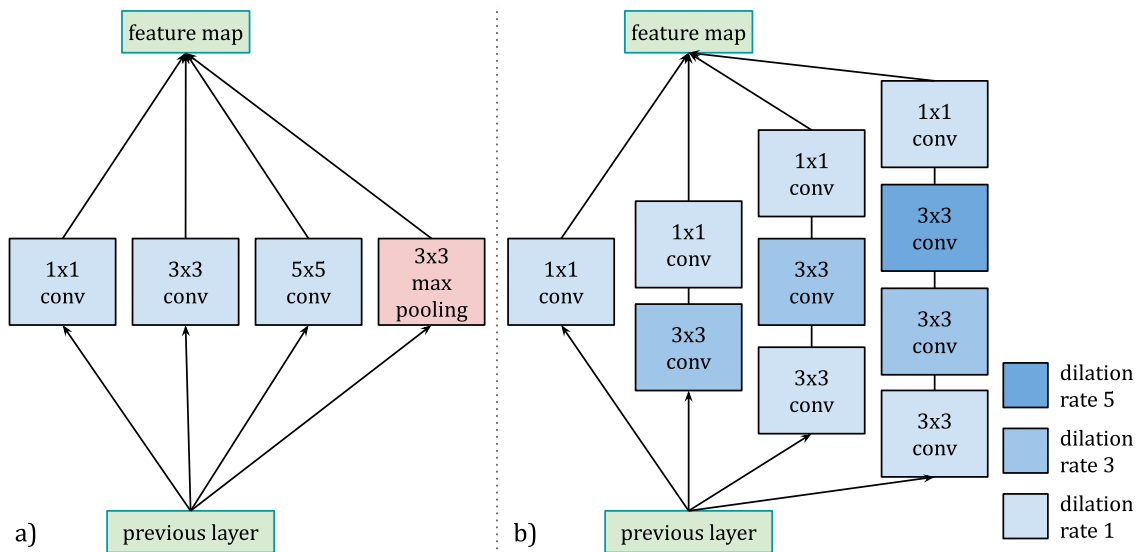


Figure 3.29: Inception module examples showing both a ‘naive’ version [285] (in a), as well as a more complex version [286] (in b). The input layer is passed through different convolutions (with different filter sizes, as well as a max-pooling operation in the ‘naive’ case, or with different dilation rates in the more complex case), and the output of each of these layers is concatenated and passed to the next layer in the network. Image adapted from [285, 286].

Qamar *et al.* [287] propose a 3D U-Net in which they adopt the inception module, as well as dense and residual connections for iso-intense MRI infant brain tissue segmentation [283]. Gu *et al.* [286] introduce a modified inception module (Figure 3.29 b) into their proposed medical image segmentation architecture, which included atrous (dilated) convolutions with the aim of widening the receptive field. The authors applied their model on multiple tasks including CT lung segmentation, retinal vessel detection, as well as cell contour segmentation, and achieved superior results to other state-of-the-art models. However, their work was focus on 2D images only, and it is also worth mentioning that the inception structure is generally quite complex and leads to increased efforts when trying to change the model’s architecture.

Loss functions

Besides developing new and improved network architectures or blocks, designing loss functions has also been of particular interest in the literature [262]. This section therefore focuses on the most prevalent loss functions designed for medical image segmentation tasks.

Cross entropy. One of the most popular loss functions is the cross entropy (CE) loss, which measures how well the estimated class probabilities match the target values. Let $K \in \mathbb{N}_{\geq 2}$ be the number of classes, y_k the value for the target class k , and \hat{p}_k be the prediction for class k , then the CE for instance (voxel) i is defined as:

$$\text{CE}^{(i)} = - \sum_{k=1}^K y_k^{(i)} \log \left(\hat{p}_k^{(i)} \right) \quad (3.13)$$

where $y_k^{(i)}$ represents the target (ground truth) probability values for each class k such that $\sum_k y_k^{(i)} = 1$. In practice, $y^{(i)}$ is often a one-hot encoded vector. Moreover, during neural network training, the loss is often calculated as an average⁴ across all N voxels:

$$\mathcal{L}_{CE} = - \frac{1}{N} \sum_{i=1}^N \left(\sum_{k=1}^K y_k^{(i)} \log \left(\hat{p}_k^{(i)} \right) \right) \quad (3.14)$$

When $K = 2$, CE becomes the binary cross entropy (BCE):

$$\text{BCE}^{(i)} = - \left(y^{(i)} \log \left(\hat{p}^{(i)} \right) + (1 - y^{(i)}) \log \left(1 - \hat{p}^{(i)} \right) \right) \quad (3.15)$$

where $y = y_1$ and $\hat{p} = \hat{p}_1$, and as the two classes are mutually exclusive we can also write: $y_2 = 1 - y$ and $\hat{p}_2 = 1 - \hat{p}$. Figure 3.30 shows how the BCE behaves across different predicted \hat{p} values for the two target classes (when $y = 1$ or when $y = 0$).

Moreover, when data imbalance is an issue, one can use the weighted cross entropy (WCE) introduced by Long *et al.* [288]:

$$\text{WCE}^{(i)} = - \sum_{k=1}^K w_k y_k^{(i)} \log \left(\hat{p}_k^{(i)} \right) \quad (3.16)$$

where w_k is a per-class weight.

In the medical imaging field, Zhang *et al.* [178] used the CE loss to train their proposed 2D CNN solution for brain tissue segmentation of 6–8 month old infants. More specifically, using patches of multi-modality information (T_1w , T_2w and FA maps) as input, their network was trained to predict the label for each individual

⁴Machine learning software frameworks such as PyTorch allow the user to choose between different reduction schemes: sum or average over all voxels.

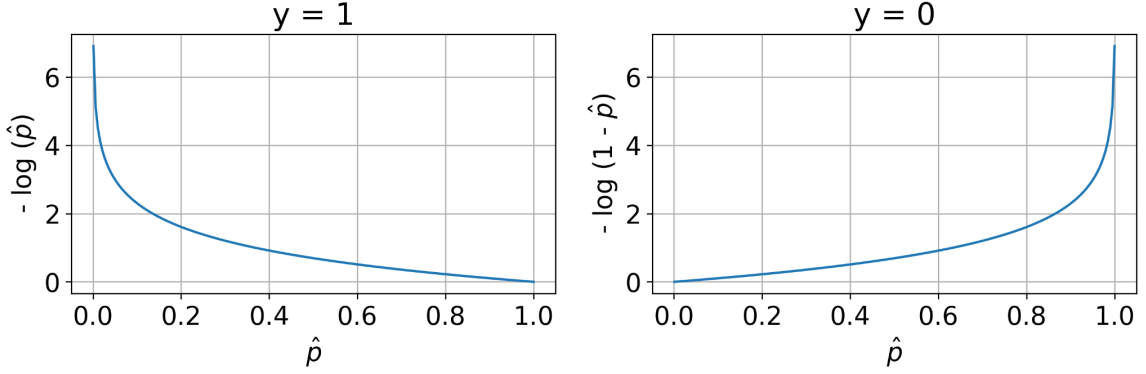


Figure 3.30: Binary cross entropy for the two target classes: $y = 1$ (left panel) and $y = 0$ (right panel). BCE becomes $-\log(\hat{p})$ when $y = 1$, and $-\log(1 - \hat{p})$ when $y = 0$, respectively. The loss penalizes the predictions more when they are further away from the target value.

voxel. Moeskops *et al.* [289] built an improved 2D CNN architecture using multi-scale information: for each voxel to be predicted, their network used patches of different sizes, as well as different sized convolution kernels. Similar to Zhang *et al.* [178], the authors trained their model with CE loss, but applied it to multiple datasets: T_2w coronal and axial slices of preterm infants (acquired at 30 and 40 weeks PMA), as well as T_1w images of an adult cohort. Moreover, they evaluated their technique on the NeoBrainS12 challenge data [164] and obtained accurate segmentations in terms of Dice scores for all tissue classes.

Ronneberger *et al.* [205] notably introduced a version of the WCE loss where the weight was only attributed to the foreground class. More specifically, w was a pixel-wise weight map, pre-calculated for each ground truth segmentation with the aim of improving predictions in areas where the separation border between two or more foreground classes was very small.

Dice loss. The DSC (see equation 2.42) is a highly regarded evaluation metric for medical imaging segmentation tasks. Milletari *et al.* [197] introduced it as the Dice loss (DL) function, where its multi-class variant is defined as:

$$\mathcal{L}_{DL} = 1 - 2 \frac{\sum_{k=1}^K \sum_{i=1}^N y_k^{(i)} \hat{p}_k^{(i)} + \epsilon}{\sum_{k=1}^K \left(\sum_{i=1}^N y_k^{(i)} \sum_{i=1}^N \hat{p}_k^{(i)} \right) + \epsilon} \quad (3.17)$$

Here, the ϵ term is added to ensure the stability of the loss during training and avoid division by 0 when the ground truth and predicted segmentations are empty.

Sudre *et al.* [290] introduced the generalised Dice loss (GDL) as a means of counteracting inter-class imbalance. The aim is to correct the contribution of each label by weighting it with the squared inverse of its volume, thus ensuring that smaller sized objects are contributing to the overall loss and are not overpowered by

larger entities. Mathematically, the GDL is defined as:

$$\mathcal{L}_{GDL} = 1 - 2 \frac{\sum_{k=1}^K w_k \sum_{i=1}^N y_k^{(i)} \hat{p}_k^{(i)} + \epsilon}{\sum_{k=1}^K w_k \left(\sum_{i=1}^N y_k^{(i)} \sum_{i=1}^N \hat{p}_k^{(i)} \right) + \epsilon} \quad (3.18)$$

where

$$w_k = \frac{1}{\left(\sum_{i=1}^N y_k^{(i)} \right)^2}$$

Loss improvements. In current state-of-the-art biomedical image segmentation literature many authors adopt the use of both CE and DL in their final loss function [266, 268, 291]. Salehi *et al.* [292] introduced the Tversky loss, based on the Tversky index [293]:

$$\text{Tversky} = \frac{|S_g^f \cap S_p^f|}{|S_g^f \cap S_p^f| + \alpha |S_p^f \setminus S_g^f| + \beta |S_g^f \setminus S_p^f|} \quad (3.19)$$

which is a generalisation of the DSC (it becomes DSC when $\alpha = \beta = 0.5$). The 2 parameters, α and β , control the amount of penalising FPs and FNs in the loss function. The authors applied their method on T_1w , T_2w and fluid attenuated inversion recovery (FLAIR) MRI volumes with the aim of segmenting multiple sclerosis lesions, and showed that they outperformed networks trained using classic Dice. Other works include the Wasserstein [294] or the focal loss [295] for imbalanced segmentation tasks, introduce penalties in the loss function through distance maps [296], or anatomically constrain predicted segmentations through a convolutional AE trained with ground truth labels [297].

3.2.3 Visual attention

Attention in the context of deep learning has become an important area of research in recent years as it can be easily incorporated in current neural network architectures, while also improving their performance [298, 299]. Attention has reached many areas of deep learning, and it is used with pixels in an image, words in a sentence [300], nodes in a graph [301], or even points in a 3D point cloud [302]. Attention was born in the area of sequence to sequence modeling [303], where networks are trained to translate sentences of arbitrary length from one language (*i.e.*, English) to another (*i.e.*, French). In this case, attention helps with figuring out the most important elements in the input sequence to predicting an accurate output sentence [304]. This dependency is also important in computer vision tasks, where attention can be used along the spatial or channel domain.

In the medical image analysis field, attention has become an important area of research as models which incorporate it attain state-of-the-art results, as well as explainability [305]. The latter is especially desirable when applied to medical

diagnosis [305, 306] where the reliance of the networks on the correct features must be guaranteed [306, 307]. For the purpose of this thesis, the focus is on image-based methods only, where attention is predominantly built as a *mask* used to identify key features in the input data or the feature maps of the neural network. In this respect, attention methods can be divided in two categories, **soft attention** and **hard attention**, depending on how the *mask* is constructed. In the **hard** case, the model is restricted to use only a subset of the input data (*e.g.*, train a network to localise an organ of interest and mask the original image using the prediction [308]), while in the **soft** case, the model pays higher or lower importance over different areas of the input. The remainder of this section will focus on discussing **soft attention** models in more depth and introduce typical examples, while **hard attention** is out of scope for this thesis.

Soft attention methods rely on building a probability map over input data, features or channels. They often add computational complexity to an existing deep learning model, but have a differentiable objective and are easily trainable with gradient descent. Soft attention can be further divided into: channel attention, spatial attention, mixed attention, and non-local attention.

Channel attention. In a typical CNN the convolutional layer weighs each of its channels equally when creating the output feature maps. Given an input tensor \mathbf{X}' with dimensions $H' \times W' \times C'$ (height, width and number of channels, respectively), the output of applying a convolutional layer is a feature map of dimensions $H \times W \times C$. Internally, this layer applies C filters of dimension $k \times k \times C'$, where k is the chosen kernel size. The purpose of channel attention is therefore to assign weights to each of the C filters in order to emphasize useful features.

One of the most popular channel attention blocks, known as the *squeeze-and-excitation* block, was introduced by Hu *et al.* [309]. Figure 3.31 shows a schematic representation of this module. For the sake of completion, the first step in this diagram shows a generic convolutional layer (\mathbf{F}_{tr}) transforming the input tensor \mathbf{X}' into a feature map \mathbf{U} with dimensions $H \times W \times C$. This feature map is then put through a *squeeze* function (\mathbf{F}_{sq}), which is defined as a global average pooling operation [310]. For each channel $c \in [1, C]$ this can be written as:

$$\mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (3.20)$$

where \mathbf{u}_c is the c^{th} element of \mathbf{U} . Its output z is then passed to an *excitation* function (\mathbf{F}_{ex}) defined as a sigmoid function applied to an MLP g with weights \mathbf{W} :

$$\mathbf{F}_{ex}(z, \mathbf{W}) = \sigma(g(z, \mathbf{W})) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 z)) \quad (3.21)$$

Here, δ is the ReLU function, σ is the sigmoid activation function, while $\mathbf{W}_1 \in \mathbb{R}^{C/2 \times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times C/2}$ are 2 fully connected layers. This operation generates a feature weight map of values constrained between $[0, 1]$ which is used to scale the input \mathbf{U} channel-wise to become the refined output \mathbf{X} . In short, the channel

attention is computed as:

$$\alpha_{SE} = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \text{GlobalAvgPool}(\mathbf{U}))) \quad (3.22)$$

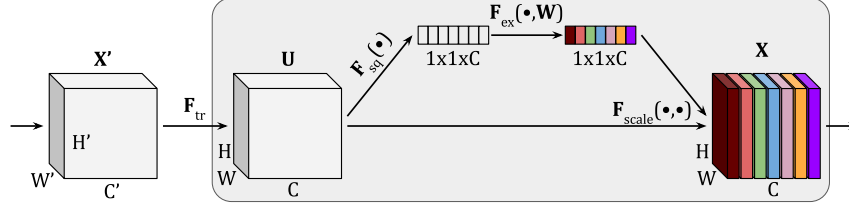


Figure 3.31: *Squeeze-and-excitation* [309] block showing: \mathbf{F}_{tr} - a generic convolutional layer which transforms the input tensor \mathbf{X}' into a feature map \mathbf{U} , \mathbf{F}_{sq} - the squeeze function applied to \mathbf{U} which produces a tensor of dimension $1 \times 1 \times C$; \mathbf{F}_{ex} - the excitation function applied to the output of the previous layer which is a MLP with a sigmoid activation function; and \mathbf{F}_{scale} - a scaling function which performs elementwise multiplication (along the channel dimension) between the input \mathbf{U} and the previous layer. The output of the *squeeze-and-excitation* module is a refined tensor \mathbf{X} of the same dimensions as its input \mathbf{U} . Image adapted from [309].

Hu *et al.* [309] introduced the *squeeze-and-excitation* block throughout inception networks [285] and ResNet [311], and showed improved classification performance on the ImageNet dataset [312]. In the medical imaging field, Chen *et al.* [313] proposed a modified U-Net [205] with ResNet blocks [311] in the encoder, and *squeeze-and-excitation* blocks to achieve feature channel attention when fusing the encoder with the decoder features. The authors used the proposed method for liver lesion segmentation in CT slices, and achieved state-of-the-art results when compared to previous methods.

Woo *et al.* [314] introduced global max pooling on top of the global average pooling layer to generate two C -dimensional descriptors (see Figure 3.32). These feature vectors are summed and then put through a sigmoid activation function which generates the final channel attention map α_C . The map is then used to scale the input \mathbf{U} to become the refined output \mathbf{X} . In short, the channel attention is computed as:

$$\alpha_C = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \text{GlobalAvgPool}(\mathbf{U})) + \mathbf{W}_2 \delta(\mathbf{W}_1 \text{GlobalMaxPool}(\mathbf{U}))) \quad (3.23)$$

where δ , σ , \mathbf{W}_1 and \mathbf{W}_2 have the same definitions as the *squeeze-and-excitation* block [309].

Spatial attention. The spatial attention is an alternative approach to soft attention which aims to extract important information in the image domain, or across the spatial domain of a feature map. In Woo *et al.* [314], the spatial attention block first performs an average and a max pooling operations across the channels on the input \mathbf{U} , generating two feature maps which are concatenated. A 7×7 convolutional layer [314] is then applied to produce a 1-channel spatial map which, after passing through a sigmoid activation function, becomes the attention map.

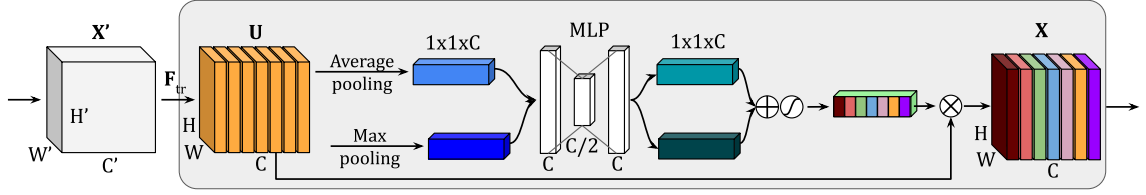


Figure 3.32: Channel attention block [314] showing: \mathbf{F}_{tr} - a generic convolutional layer which transforms the input tensor \mathbf{X}' into a feature map \mathbf{U} , followed by global average and max pooling layers. The two feature descriptors are forwarded to a shared network (an MLP with one hidden layer) to produce refined feature maps. These two vectors are merged using element-wise summation and put through a sigmoid activation function to create the final channel attention map. The output \mathbf{X} is generated by multiplying the attention map by the original input \mathbf{U} . Image adapted from [314].

The input \mathbf{U} is then scaled by the attention map becoming the refined feature map \mathbf{X} . In short, the spatial attention is computed as:

$$\alpha_{\mathbf{S}} = \sigma \left(f^{7 \times 7} ([\text{AvgPool}(\mathbf{U}); \text{MaxPool}(\mathbf{U})]) \right) \quad (3.24)$$

where $[\cdot; \cdot]$ represents the channel-wise concatenation of the two feature maps and $f^{7 \times 7}$ is the convolutional layer. Figure 3.33 shows a schematic representation of this module.

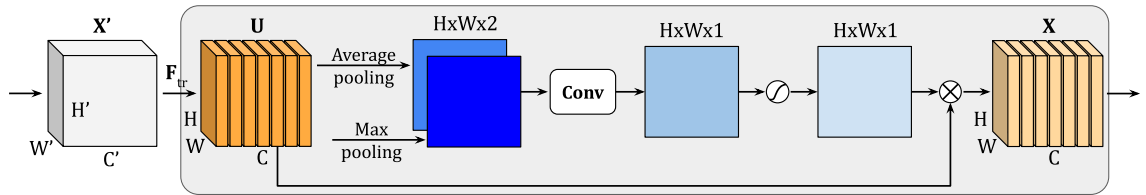


Figure 3.33: Spatial attention block [314] showing: \mathbf{F}_{tr} - a generic convolutional layer which transforms the input tensor \mathbf{X}' into a feature map \mathbf{U} , followed by average and max pooling layers. Then, a 7×7 convolutional layer is applied on the concatenated maps, while the sigmoid activation function creates the attention map. The output \mathbf{X} is generated by multiplying the attention map by the original input \mathbf{U} . Image adapted from [314].

In the medical imaging field, Guo *et al.* [315] introduced the spatial attention block [314] in the bottleneck of a 2D U-Net [205] and showed that their proposed method outperformed the classic U-Net in terms of segmenting fine structures (blood vessels) in retinal fundus image data. Similarly, Oktay *et al.* [265] introduce the Attention U-Net, where attention gates are added to every decoding layer. Unlike the spatial attention block [314] shown in Figure 3.33 where the average and max pooling layers are applied to the same input feature map, the attention gates apply convolutions to both features from the encoder and the corresponding decoder and fuse them together before creating the attention map. Moreover, instead of simply concatenating the encoder feature maps through the use of skip connections, the authors first scale them with the generated spatial attention. The proposed Attention U-Net was evaluated on 3D multi-class abdominal CT segmentation where it

showed improved results against standard U-Net, especially in organs with variable small size.

Mixed attention. Woo *et al.* [314] also introduced an aggregated attention with the aim of combining the advantages of both spatial and channel attention mechanisms. The authors proposed the convolutional block attention module (CBAM), where the channel attention block (Figure 3.32) is followed by the spatial attention block (Figure 3.33) as shown in Figure 3.34.

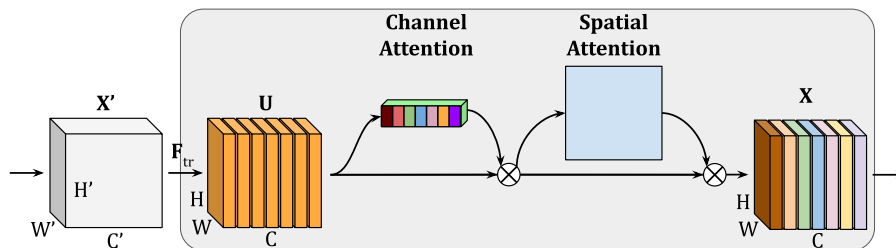


Figure 3.34: Mixed channel and spatial attention block (CBAM) [314] where the feature map \mathbf{U} is first refined using channel attention (Figure 3.32) and then using spatial attention (Figure 3.33), to produce the final output \mathbf{X} . Image adapted from [314].

In the medical imaging field, CBAM [314] has proven to be a popular option for integrating attention into existing CNN architectures. For example, Zhao *et al.* [316] introduced the CBAM U-Net++ medical image segmentation network, which, as the name suggests, combines U-Net++ [264, 281] with mixed channel and spatial attention. The authors showed improved performance on nuclei segmentation of biological images when compared to both classic U-Net [205] and U-Net++. Similarly, CBAM was added to the respective neural network architecture for segmenting the sclera in 2D images [317], the hippocampus in 3D MR images [318], or polyps in 2D colonoscopy videos [319], among others.

Non-local attention. Despite their ability to improve the final segmentation performance, channel and spatial attention mechanisms focus mainly on local information. In both cases, the operation of max or average pooling leads to loss of spatial information, while convolutional layers process neighbourhood information. To overcome such limitations, Wang *et al.* [320] proposed non-local attention, which aims to capture long-range dependencies by computing interactions between any two positions in an image or feature map, with a better awareness of the entire context. Their strategy follows closely the ideas brought forward by the self-attention mechanism [300] introduced for natural language processing.

The overall architecture of the non-local attention block is shown in Figure 3.35, where the authors chose to embed it in a neural network model through the use of residual connections (*i.e.*, through addition). As an initial step, Wang *et al.* [320] proposed three parallel 1^2 convolutional operations (θ , φ and g) to be applied on the input \mathbf{U} , obtaining three compressed feature maps. Introduced by Lin *et al.* [310], the $1 \times 1 (\times 1)$ convolutions (also known as projection layers) are often used

for dimensionality reduction (in this case the feature maps go from C to $C/2$), and act as a channel-wise pooling operator.

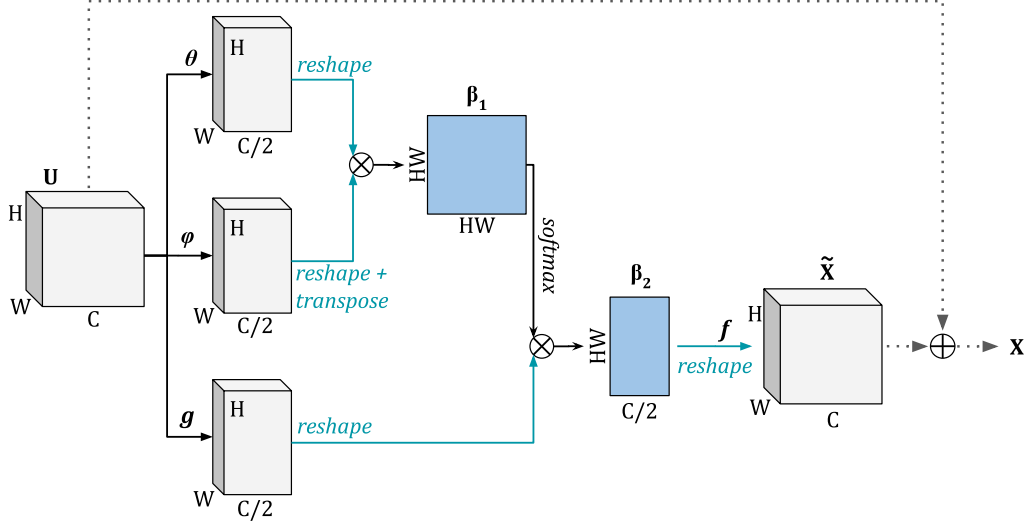


Figure 3.35: Non-local attention [320] where the feature map \mathbf{U} is first transformed by a series of parallel 1^2 convolutional layers (θ, φ, g). β_1 is generated through matrix multiplication between $\theta(\mathbf{U})$ and $\varphi(\mathbf{U})$, while β_2 is generated by multiplying β_1 with $g(\mathbf{U})$. The outputs of several steps are reshaped accordingly to allow for matrix multiplications. Note that \otimes represents the matrix multiplication operator, while \oplus is the element-wise addition. Image adapted from [320].

The individual maps are then reshaped into 2D matrices ($HW \times C/2$), and an initial feature map is calculated:

$$\beta_1 = \text{reshape}(\theta(\mathbf{U})) \otimes \text{reshape}(\varphi(\mathbf{U}))^T \quad (3.25)$$

where $\text{reshape}(\cdot)$ changes the tensor to be in the $HW \times C/2$ configuration, T is the matrix transpose operation and \otimes is the matrix multiplication operator.

The next step is achieved through a second matrix multiplication between β_1 and the output of the convolutional layer g :

$$\beta_2 = \sigma(\beta_1) \otimes \text{reshape}(g(\mathbf{U})) \quad (3.26)$$

The last step is to reshape the matrix back into a tensor of dimensions $H \times W \times C/2$, apply a final 1^2 convolutional layer with C output channels:

$$\tilde{\mathbf{X}} = f(\text{reshape}(\beta_2)) \quad (3.27)$$

and generate the output through element-wise addition: $\mathbf{X} = \mathbf{U} + \tilde{\mathbf{X}}$.

In the medical imaging field, Wang *et al.* [321] proposed a 3D non-local U-Net and showed that through the use of attention blocks they were able to improve segmentation accuracy of tissue maps in isointense infant brain MR volumes. Similarly, Gu *et al.* [322] introduced the non-local attention block in a modified U-Net, but also added channel and spatial attention blocks throughout the decoder layers of

the network. The authors applied it to skin lesion segmentation and multi-class segmentation of fetal MRI, and showed that for the latter their proposed architecture achieves higher accuracy when compared to previous state-of-the-art segmentation models.

3.2.4 Domain adaptation

Deep learning models suffer from the *domain shift* [323] problem, which refers to the difference in data distributions between training and testing datasets. This is prevalent in the medical community where heterogeneity in the data arises from multi-center studies, different acquisition protocols, patient biases, and imaging modalities. DA aims to tackle this issue by minimizing the gap between the two (or more) domains, as long as they share the same learning tasks. This is exemplified in Figure 3.36 where a classifier has been trained on the source domain (A), but performs poorly on the target domain (B). After performing DA, the classifier can now successfully be used on both datasets.

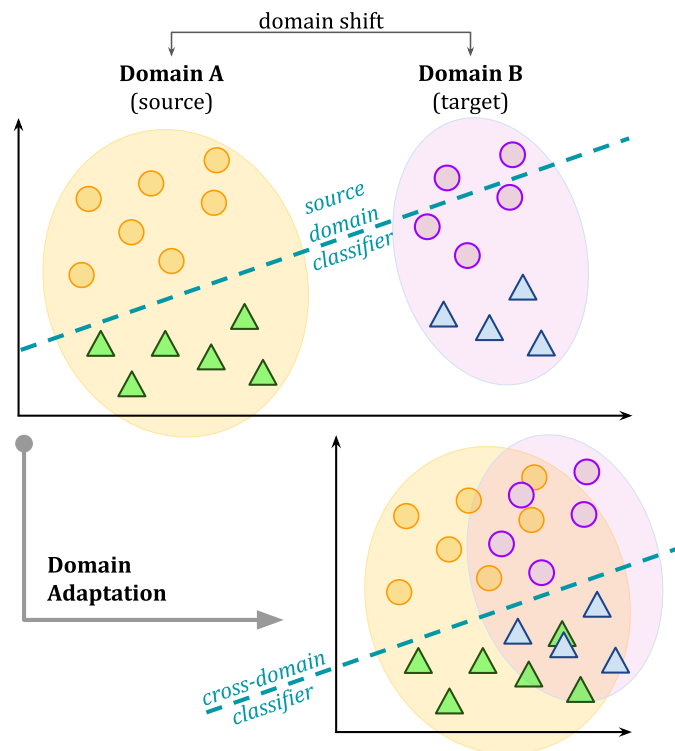


Figure 3.36: Domain adaptation aims to solve the *domain shift* problem commonly found in many machine learning algorithms where a model trained on domain A cannot be reliably used on a second domain, B, as exemplified in the top half of the figure. Through domain adaptation (bottom half of the image) the domain shift has been corrected and the classifier can be used on both datasets.

In medical imaging, this issue is prevalent especially in MRI acquisitions, as the images are not quantitative [324]. In fact, both inter- and intra-scanner variability exists [325, 326], which makes DA especially important for downstream MRI

analysis. Figure 3.37 illustrates the *domain shift* issue in terms of intensity distributions of structural T_2w MR images from 2 datasets: dHCP [11] and ePrime [35]. Conventional machine learning techniques generally ignore these problems, which consequently degrades their performance [327]. To solve this, DA has recently become an important topic of medical imaging research [328, 329, 330].

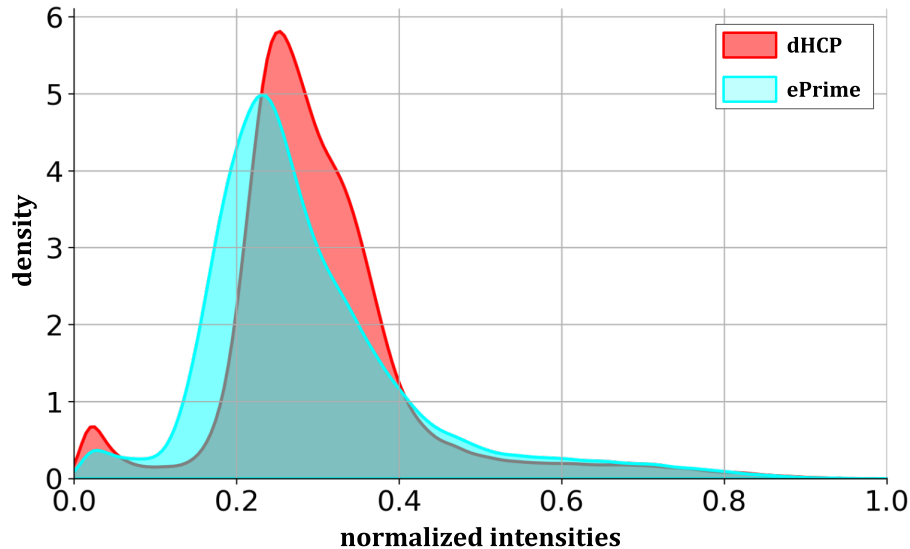


Figure 3.37: Intensity distribution of T_2w MR images from the dHCP and ePrime datasets. Intensity is normalized between 0 and 1.

In this section, the focus is on reviewing DA techniques applied to medical imaging data. For this purpose, the different methods presented here are categorised by label availability. More specifically, DA methods can be grouped into: supervised, semi-supervised and unsupervised techniques, where in the first two cases a small number of labelled data is available when training the models. Due to their scarcity in the medical imaging field, a more useful approach is the unsupervised category which assumes unlabelled target data. For a more in-depth survey, the reader can consult Guan *et al.* [331].

Supervised domain adaptation

One approach to supervised DA is to directly transfer a pre-trained model to the target domain. For example, Ghafoorian *et al.* [329] first trained a segmentation network for brain white matter hyperintensities on a source domain of MRI scans and then performed fine-tuning on the target domain. More specifically, they investigated the impact of freezing different number of layers in the architecture and training the remaining ones on a target domain dataset. Moreover, the authors evaluated the influence the number of target labels has on the performance of the segmentation network. The authors showed that the domain adapted network fine-tuned on only 2 target images outperformed the same NN architecture trained on the same examples from scratch.

Gu *et al.* [332] introduce a multi-step DA approach where they first train a CNN on the ImageNet dataset [333]. Then, they fine-tune the network on a large (intermediate) medical image dataset for skin cancer, and finally, train it on the target domain which is a relatively small skin medical image dataset. Their experiments show that the multi-step approach achieves better performance than single step transfer learning.

Although popular, the aforementioned methods rely on pre-training the networks on 2D datasets (such as ImageNet [333]). This approach is not sufficient when exploring the rich information provided by the 3D datasets available in the medical imaging community. For this reason, Hosseini-Asl *et al.* [334] propose a 3D CNN for Alzheimers disease classification based on brain MR volumes. The authors first train a convolutional AE to reconstruct the source domain 3D images. Then, they freeze the network weights and attach fully-connected layers which are fine-tuned with samples from the target data. A similar approach is proposed by Valverde *et al.* [335] for multiple sclerosis brain MR image segmentation, where they investigated the number of fully-connected layers to be fine-tuned using the target data and evaluated their proposed framework on the ISBI2015 dataset [336].

Semi-supervised domain adaptation

In semi-supervised DA, a small number of target labelled data as well as target unlabelled data are used to fine-tune the model. For example, Roels *et al.* [337] propose a segmentation network called Y-Net with one encoder and two decoders. The authors train one of the branches in an unsupervised way, *i.e.*, one of the decoders is trained to reconstruct both source and target images. Then, this decoder is discarded and the network is fine-tuned with labelled target data.

Liu *et al.* [338] propose a semi-supervised domain adaptation model which aims to effectively balance the larger volume of source domain labels with the much smaller amount of labelled target data, by developing an asymmetric co-training strategy. More specifically, the authors propose the co-training of 2 segmentation networks: one with labelled source domain data and unlabelled target domain data, and the other with labelled and unlabelled target domain data only. Moreover, they show that their proposed method outperforms two other state-of-the-art unsupervised DA models. However, it is important to note that both supervised and semi-supervised DA approaches require labelled target data. As labeling is a time-consuming and potentially variable effort, especially in medical imaging where there is a need for highly specialised experts, unsupervised DA methods, where the target domain does not have labels, has recently attracted more attention.

Unsupervised domain adaptation

This section introduces different unsupervised DA techniques grouped by their specific knowledge transfer strategies.

Feature alignment (adversarial domain adaptation in the latent space).

Most unsupervised DA techniques which employ the feature alignment strategy rely on the work of Ganin *et al.* [339]. Their proposed method is schematically shown in Figure 3.38. The main idea is to attach a classifier to the main network to force it to learn domain-invariant features through adversarial training. Kamnitsas *et al.* [330] introduce this method for domain adaptation of brain lesion segmentation in 3D MR volumes. More specifically, they extend their proposed DeepMedic [270] image segmentation network with a domain classifier attached to different layers of the segmentor. Through experiments they show the efficiency of their unsupervised DA framework, while also investigating which layers should be connected to the domain classifier.

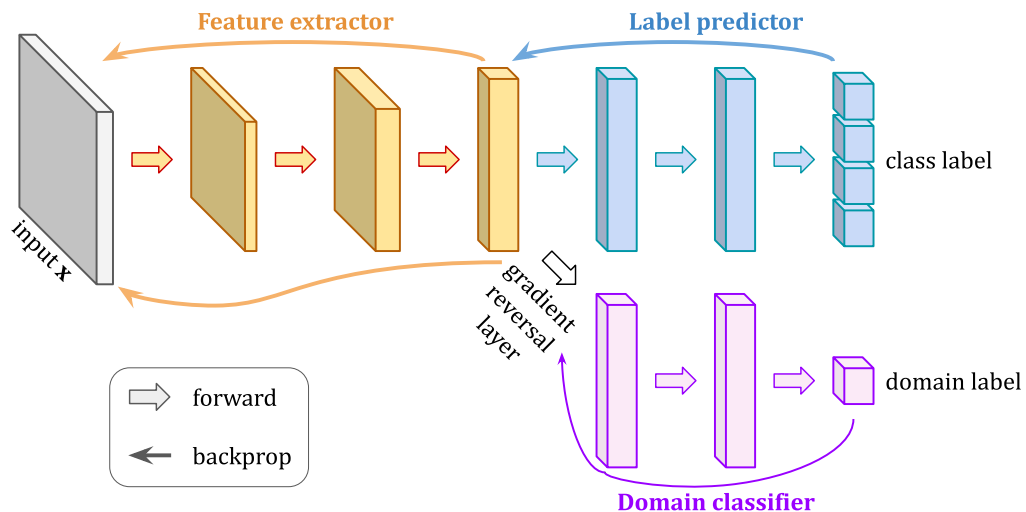


Figure 3.38: Schematic illustration of adversarial domain adaptation in the latent space. The feature extractor (yellow layers) is trained to predict a class label (blue layers), while the domain classifier (purple layers) force it to learn domain-invariant features through adversarial training. Image adapted from [339].

Dou *et al.* [340] propose an MR to CT (cross-modality) unsupervised DA model for cardiac segmentation. The authors adapt only the early encoder layers, while keeping the higher layers fixed between the two domains. Moreover, they introduce two discriminators, one for feature discrimination, and one for the predicted segmentation masks, and validate their framework on the multi-modality whole heart segmentation (MM-WHS) dataset [341].

A similar approach was proposed by Yan *et al.* [342] for cross-vendor 2D MR image segmentation. The novelty the authors bring is in introducing Canny edge [343] maps as input and at the last two layers of the segmentation network. The authors showcase their method’s performance on data from 3 independent vendors

(Philips, Siemens and GE). Bateson *et al.* [344] propose the use of a constrained prediction when performing unsupervised DA. More specifically, they introduce a network module which predicts the size of the target domain region as a regularizer to the overall architecture.

Image alignment (adversarial domain adaptation in the image space).

Besides feature alignment, image alignment is also a popular approach to unsupervised DA [331]. The key idea is to train the main predictor (*e.g.*, a segmentation network) on images synthesized from the source to the target domain. As labels are available in the source domain, the predictor can be trained in a supervised fashion. This is schematically shown in Figure 3.39 where $\mathbf{f}_{\mathbf{S} \rightarrow \mathbf{T}}$ transforms the input \mathbf{X} from the source domain \mathbf{S} to the target domain \mathbf{T} . The predictor will be trained with fake images for which source domain labels exist. At inference, $\mathbf{f}_{\mathbf{S} \rightarrow \mathbf{T}}$ can be discarded and the main predictor can then be used directly on target data.

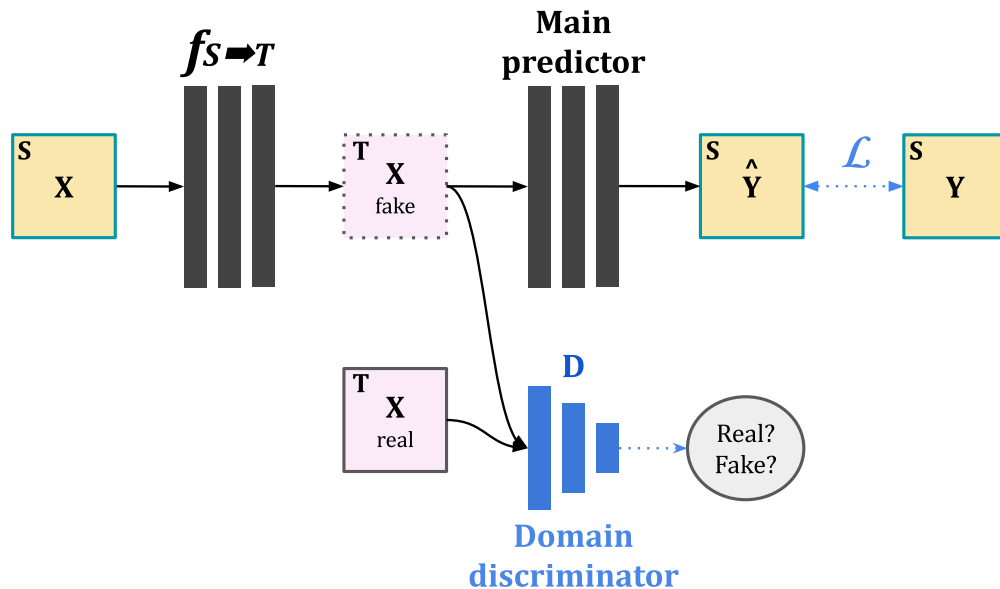


Figure 3.39: Adversarial domain adaptation in the image space where the source domain image (X^S) is first transformed with the $\mathbf{f}_{\mathbf{S} \rightarrow \mathbf{T}}$ to look like a target domain image (X^T). The main predictor (*e.g.*, a segmentation network) is trained on the fake images with source domain labels (Y^S). The discriminator is trained in a GAN-like setup [217, 220, 221] to enforce the synthesized images to look realistic.

Gholami *et al.* [345] propose the use of a Cycle-GAN [221] to generate MR images in order to augment their training dataset. As their main aim is to perform brain tumor segmentation, the authors first simulate tumor-ridden images and use the Cycle-GAN framework to make them look more realistic. Zhang *et al.* [346] introduce a noise style transfer unsupervised DA method using an image-to-image translation [220] framework, but with two discriminators. Their goal is to preserve the underlying content of the images, while transferring the noise style. Therefore, they use one of the discriminators to enforce content preservation in the generated images, and the second discriminator to enforce the same noise patterns between source and target domains. Experiments on optical coherence tomography blood

vessel segmentation showcase their method’s effectiveness.

Mahmood *et al.* [225] reverse the problem by training a GAN-like architecture to generate synthetic images from real ones. Their aim was to remove patient specific texture and details, while preserving useful diagnostic information, for depth estimation applications. The authors show that their proposed unsupervised image space DA method which transfers images from the real domain to the synthetic-like domain, managed to improve the task of endoscopy depth estimation applied to real colon data.

Li *et al.* [347] introduce a 2D neonatal brain MR image segmentation framework. Their method is trained in three stages: first, the segmentation network is trained on source data only; second, the generator is trained to perform image-level domain transfer while the segmentation network acts as a pre-trained controller to provide shape constraints; and third, the segmentation network is further trained with the synthesized images only. The authors mention that stages 2 and 3 can be repeated to further improve results. Experiments on both NeoBrainS12 [164] and dHCP [11] show improved average Dice scores compared with other state-of-the-art models. Finally, Chen *et al.* [348] propose a similar approach for neonatal brain MR image segmentation, but applied to 3D data. More specifically, their framework consists of 2 steps: first, a segmentation network is trained on source domain data only, and second, a Cycle-GAN [221] architecture is trained to perform image-to-image translation of each target data into the source domain. The authors train the two networks separately and then apply the segmentation network on synthesized target data. Through experiments they show that when compared to 2D unsupervised DA methods their proposed framework achieves improved results.

Image and feature alignment Chen *et al.* [349] use both image and feature alignment for cross-modality (MR and CT) 2D medical image segmentation. Their setup includes a Cycle-GAN [221] which transfers source domain images into the target domain, as well as a feature alignment module. The latter consists of a CNN trained on both real and synthesized images through a domain discriminator which aims to further reduce the domain shift. The authors validate their proposed method on both cardiac [341] and abdominal [350] datasets. Yan *et al.* [351] propose a similar approach for segmentation of 2D cardiac cine MR images. The authors train a U-Net [205] for image segmentation on fake, Cycle-GAN [221] generated, target domain data. Moreover, they introduce a penalty between the U-Net encoder’s feature maps of the original and the translated images with the aim of enforcing feature-level adaptation. Their experiments include data from three vendors, Philips, Siemens and GE, and the authors show that the segmentation network trained on one vendor can generalize well to the other ones without using labels.

Harmonised segmentation of neonatal brain MRI

Motivation

The performance of deep learning methods drops when applied to images acquired with acquisition protocols or patient cohorts different than the ones used to train the models.

Contribution

Investigated unsupervised DA methods and proposed the use of NCC loss to enforce image similarity between real and synthesised images, with the aim of predicting brain tissue segmentations of T_2w MRI data of an unseen neonatal population.

Publications

- Grigorescu, I. et al. (2021). *Harmonized Segmentation of Neonatal Brain MRI*. Frontiers in Neuroscience
doi.org/10.3389/fnins.2021.662005
- Grigorescu, I. et al. (2020). *Harmonized Segmentation of Neonatal Brain MRI: A Domain Adaptation Approach*. PIPPI 2020. LNCS (Springer)
doi.org/10.1007/978-3-030-60334-2_25

Code available at:

github.com/irinagrigorescu/udaneonatalmri

4.1 Introduction

Medical image deep learning has made incredible advances in solving a wide range of scientific problems, including tissue segmentation or image classification [352]. However, one major drawback of these methods is their applicability in a clinical setting, as many models rely on the assumption that the source and target domains are drawn from the same distribution. As a result, the efficiency of these models may drop drastically when applied to images which were acquired with acquisition protocols different than the ones used to train the models [330, 353].

At the same time, combining imaging data from multiple studies and sites is necessary to increase the sample size and thereby the statistical power of neuroimaging studies. However, one major challenge is the lack of standardization in image acquisition protocols, scanner hardware, and software. Inter-scanner variability has been demonstrated to affect measurements obtained for downstream analysis such as voxel-based morphometry [58], and lesion volumes [59]. Therefore, the purpose of harmonising MRI datasets is to make sure that the differences arising from different image acquisition protocols do not affect the analysis performed on the combined data. For example, volumetric and cortical thickness measures should only be affected by brain anatomy and not the acquisition protocol or scanners.

A class of deep learning methods called DA techniques aims to address this issue by suppressing the domain shift between the training and test distributions. In general, DA approaches are either semi-supervised, which assume the existence of labels in the target dataset, or unsupervised, which assume the target dataset has no labels. For example, a common approach is to train a model on source domain images and fine-tune it on target domain data [354, 329]. Although these methods can give good results, they can become impractical as more often than not the existence of labels in the target dataset is limited or of poor quality. Unsupervised domain adaptation techniques [355, 356] offer a solution to this problem by minimizing the disparity between a source and a target domain, without requiring the use of labelled data in the target domain.

In this work, we investigate two unsupervised domain adaptation methods with the aim of predicting brain tissue segmentations on T_2w 3D MRI volumes of an unseen preterm-born neonatal population. Our models are trained on a dataset with majority of term-born neonates and applied to a preterm-only population. Our key contributions are:

- We study the application and viability of unsupervised domain adaptation methods in terms of harmonising segmentations of two neonatal datasets.
- We propose an additional loss term in one of the methods, in order to constrain the network to more realistic reconstructions.
- We compare the two unsupervised domain adaptation methods with a fully-

supervised baseline and report our results in terms of Dice scores obtained on the test dataset.

- We validate the models by comparing tissue volumes and CT measures of harmonised data on two neonatal datasets acquired with different protocols and matched for GA at birth and PMA at scan.
- Finally, we perform an analysis comparing term and preterm-born neonates on the harmonised cortical gray matter maps and we show the importance of harmonising the data by a proof-of-principle investigation of the association between cortical thickness and a language outcome measure.

4.2 Methods

4.2.1 Data acquisition and preprocessing

The T_2w MRI data used in this study was collected as part of two independent projects: the developing Human Connectome Project (dHCP¹, approved by the National Research Ethics Committee REC: 14/Lo/1169), and the Evaluation of Preterm Imaging (ePrime², REC: 09/H0707/98) study. The dHCP neonates were scanned during natural unседated sleep at the Evelina London Childrens Hospital between 2015 and 2019. The ePrime neonates were scanned after being sedated, and no motion correction was applied [35]. Infants with major congenital malformations were excluded from both cohorts. Details about the data acquisition can be found in Section 1.2.3.

Our two datasets comprise of 403 MRI scans of infants (184 females and 219 males) born between 23 – 42 weeks GA at birth and scanned at term-equivalent age (after 37 weeks PMA) as part of the dHCP pipeline, and a dataset of 486 MRI scans of infants (245 females and 241 males) born between 23 – 33 weeks GA and scanned at term-equivalent age as part of the ePrime project. Figure 4.1 shows their age distribution.

Both datasets were pre-processed prior to being used by the deep learning algorithms. The ePrime volumes were linearly upsampled to 0.5 mm isotropic resolution to match the resolution of our source (dHCP) dataset. Both dHCP and ePrime datasets were rigidly aligned to a common 40 weeks gestational age atlas space [357] using the MIRTk [68] software toolbox. Then, skull-stripping was performed on all of our data using the brain masks obtained with the Draw-EM pipeline for automatic brain MRI segmentation of the developing neonatal brain [148]. Tissue segmentation maps were obtained using the same pipeline (Draw-EM) for both

¹<http://www.developingconnectome.org/>

²<https://www.npeu.ox.ac.uk/prumhc/eprime-mr-imaging-177>

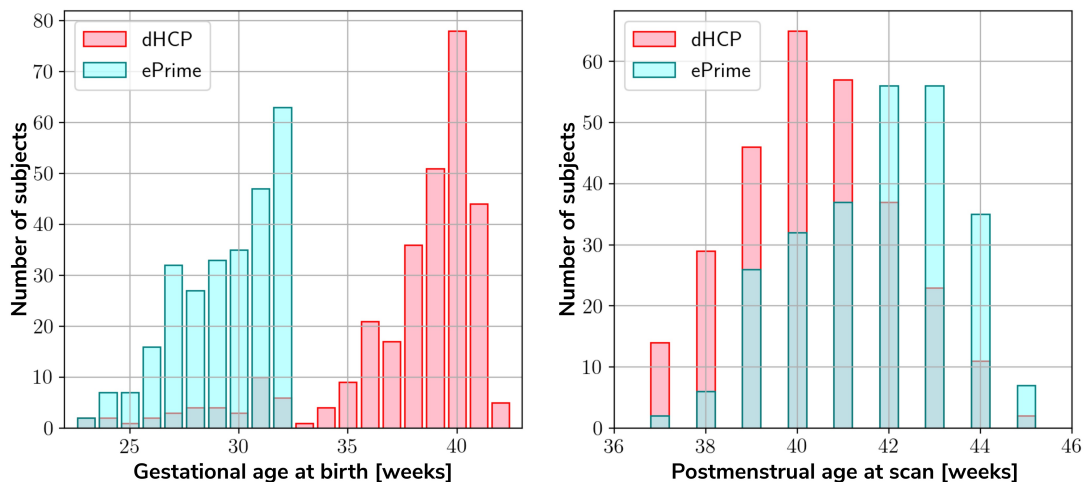


Figure 4.1: Age distribution of subjects in our dHCP and ePrime datasets, showing both their GA at birth, as well as their PMA at scan.

Dataset	#Subjects	GA [weeks]	PMA [weeks]
Train dHCP	340 (160♀ + 180♂)	39.1 (± 2.7)	40.7 (± 1.7)
Validate dHCP	32 (12♀ + 20♂)	39.3 (± 1.6)	40.7 (± 1.8)
Test dHCP	30 (12♀ + 19♂)	30 (± 2.4)	41.4 (± 1.7)
Train ePrime	417 (214♀ + 203♂)	29.6 (± 2.3)	42.9 (± 2.6)
Validate ePrime	38 (18♀ + 20♂)	29.8 (± 2.3)	43 (± 2.6)
Test ePrime	30 (13♀ + 18♂)	30 (± 2.4)	41.4 (± 1.7)

Table 4.1: Number of scans in different datasets used for training, validation and testing the models, together with their mean GA at birth and PMA at scan.

(dHCP and ePrime) cohorts, and in this study we call them: ‘original dHCP’ and ‘original ePrime’. It is worth noting here that the Draw-EM pipeline does not produce quality results on the ePrime dataset, and for this reason we do not use the predicted segmentation maps for training purposes.

In order to allow for a fair comparison between the dHCP and ePrime datasets, we first looked for a subset of neonates whose ages at birth and at scan matched. We found and selected 30 dHCP and 30 ePrime subjects with a 1-to-1 correspondence of ages at birth and at scan, and these became our test dataset (see Table 4.1). It is worth pointing out that although dHCP contains both term and preterm neonates, in the test dataset we only used preterm infants in order to match the ePrime data. To train our networks, we split the remaining data into 90% training and 10% validation (see Table 4.1), keeping both the distribution of ages at scan and the male-to-female ratio as close to the original as possible. We used the validation sets to keep track of our models’ performance during training, and the test sets to report our final models’ results and showcase their capability to generalize.

4.2.2 Unsupervised domain adaptation models

To investigate the best solution for segmenting our target dataset (ePrime), we compared three independently trained deep learning models:

- **Baseline.** A 3D U-Net [215] trained on the source dataset (dHCP) only and used as a baseline segmentation network (see Figure 4.2).
- **Adversarial domain adaptation in the latent space.** A 3D U-Net segmentation network trained on source (dHCP) volumes, coupled with a discriminator trained on both source (dHCP) and target (ePrime) datasets (see Figure 4.3). This solution is similar to the one proposed by [330] where the aim was to train the segmentation network such that it becomes agnostic to the data domain.
- **Adversarial domain adaptation in the image space.** Two 3D U-Nets, one acting as a generator, and a second one acting as a segmentation network, coupled with a discriminator trained on both real and synthesised ePrime volumes. The segmentation network is trained to produce tissue maps of the synthesised ePrime volumes created by the generator (see Figure 4.4). The NCC loss is added to the generator network to enforce image similarity between real and synthesised images, a solution which we previously proposed in [358].

To further validate the harmonised tissue maps, we trained an additional network (a 3D U-Net) to segment binary cortical tissue maps into 11 cortical substructures (see Tables 1 and 2 in Appendix A) based on anatomical groupings of cortical regions derived from the Draw-EM pipeline. The key reasons for training an extra network are: first, we avoid the time consuming task of label propagation between our available dHCP Draw-EM output segmentations and predicted ePrime maps, and second, we can train this network using Draw-EM cortical segmentations, and apply it on any brain cortical gray matter maps as in this case there will be no intensity shift between target and source distributions.

4.2.3 Network architectures

The segmentation networks in all three setups and the generator used in the image space adversarial domain adaptation model have the same architecture, consisting of 5 encoding branches with 16, 32, 64, 128 and 256 channels, respectively, and 5 decoding branches with 128, 64, 32, 16, and the number of output channels, respectively. The encoder blocks use 3^3 convolutions (with a stride of 1), instance normalisation [199] and LeakyReLU activations. A 2^3 average pooling layer is used after the first down-sampling block, while the others use 2^3 max pooling layers. The decoder blocks consist of 3^3 convolutions (with a stride of 1), instance normalisation

[199], LeakyReLU activations, and, additionally, 3^3 transposed convolutions. The number of encoding-decoding blocks, as well as the use of LeakyReLU activations and instance normalisation layers, were chosen based on the best practices described in [216]. At the same time, the network configurations that we have chosen allowed us to work with the hardware we have at hand (Titan XP 12 GB). The segmentation network outputs a 7-channel 3D volume (of the same size as the input image), corresponding to our 7 classes: background, CSF, cGM, WM, dGM, cerebellum and brainstem. The generator network’s last convolutional layer is followed by a Tanh activation and outputs a single channel image.

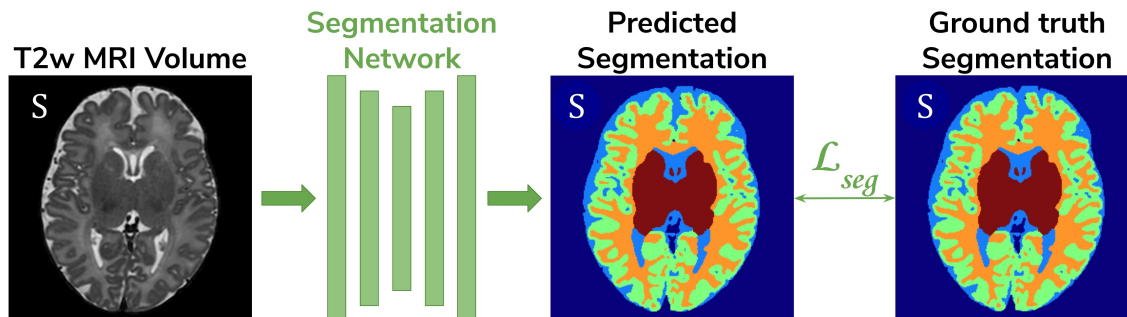


Figure 4.2: The baseline model consists of a 3D U-Net trained to segment source (dHCP) volumes. The input T_2w MRI images, the predicted segmentation and the Draw-EM output segmentations are marked with S as they all belong to the source (dHCP) dataset.

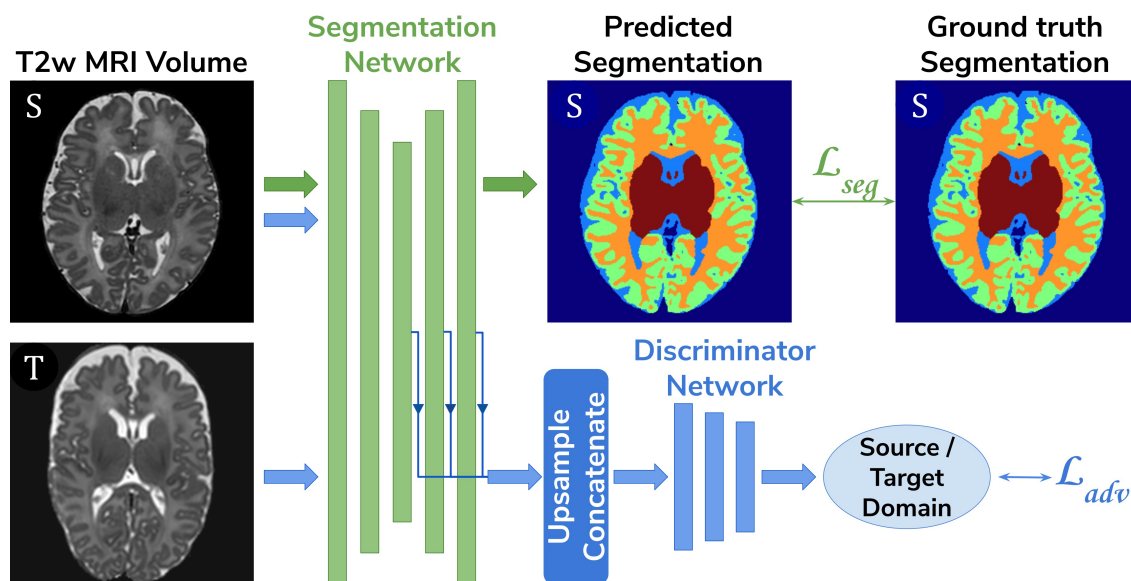


Figure 4.3: The latent space domain adaptation setup consists of a 3D U-Net trained to segment the source (dHCP) T_2w MRI volumes, coupled with a discriminator network which forces the segmentation network to learn domain-invariant features. Both source (dHCP) and target (ePrime) images are fed to the segmentation network, but only source (dHCP) Draw-EM output labels are used to compute the segmentation loss. Source domain images are marked with S, while target domain images are marked with T, respectively.

For our unsupervised domain adaptation models (Figures 4.3 and 4.4) we used a PatchGAN discriminator as proposed in [220]. Its architecture consists of 5 blocks

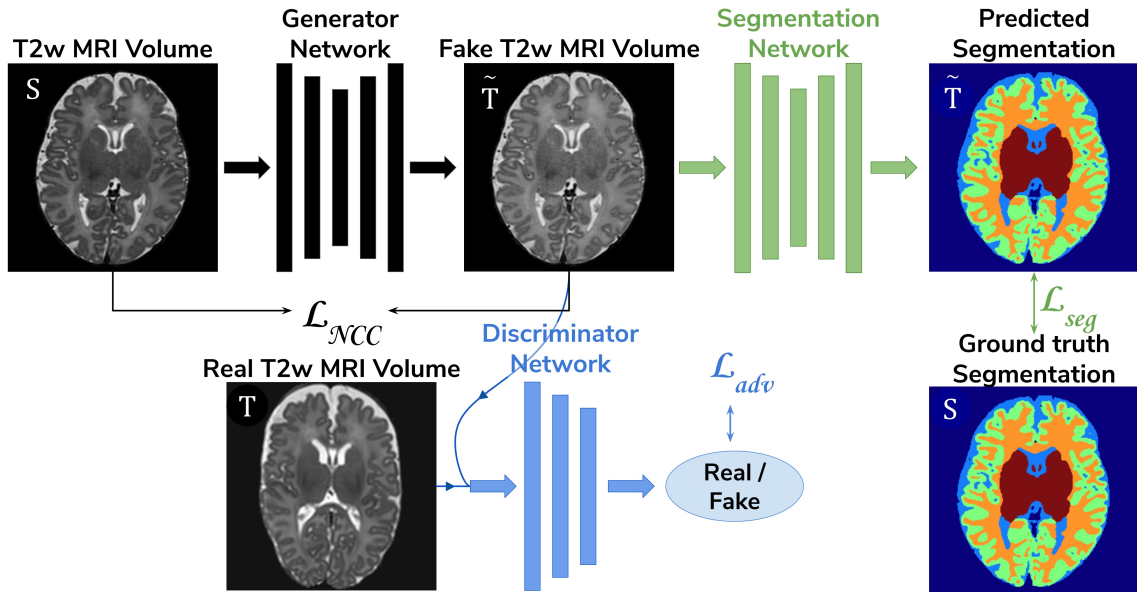


Figure 4.4: The image space domain adaptation setup uses a generator network to produce ePrime-like T_2w MRI images (marked with \tilde{T}), which are then used as input into the segmentation network. The discriminator is trained to distinguish between real (ePrime) and synthesised (ePrime-like) volumes, while the generator is trained to produce realistic images in order to fool the discriminator. The NCC loss enforces image similarity between real and synthesised volumes.

of 4^3 convolutions (with a stride of 2) with 64, 128, 256, 512 and 1 channels, respectively), instance normalisation and LeakyReLU activations.

The cortical parcellation network has the same architecture as the tissue segmentation network, but outputs a 12-channel 3D volume corresponding to the following cortical substructures: frontal left, frontal right, cingulate, temporal left, temporal right, insula left, insula right, parietal left, parietal right, occipital left, and occipital right, respectively. The last class represents the background.

4.2.4 Training

The baseline segmentation network (Figure 4.2) was trained by minimizing the generalised Dice loss [290] between the predicted and the Draw-EM segmentation maps (Equation 4.1).

$$\mathcal{L}_{method_1} = \mathcal{L}_{seg} = 1 - 2 \frac{\sum_{l=1}^M w_l \sum_n p_{ln} t_{ln}}{\sum_{l=1}^M w_l \sum_n p_{ln} + t_{ln}} \quad (4.1)$$

where $w_l = 1/(\sum_n t_{ln})^2$ is the weight of the l^{th} tissue type, p_{ln} is the predicted probabilistic map of the l^{th} tissue type at voxel n , t_{ln} is the target label map of the l^{th} tissue type at voxel n , and M is the number of tissue classes. While training, we used the Adam optimizer [201] with its default parameters and a decaying cyclical

learning rate scheduler [202] with a base learning rate of $2 \cdot 10^{-6}$ and a maximum learning rate of $2 \cdot 10^{-3}$. The choice of optimizer was based on knowledge of previous image translation literature [220, 221, 359, 360] where it yielded good results. At the same time, a varying learning rate during training was shown to improve results in fewer iterations when compared to using a fixed value [202].

The segmentation network from the adversarial domain adaptation in the latent space model was trained to produce tissue maps on the source (dHCP) volumes. In addition, both target (ePrime) and source (dHCP) volumes were fed to the segmentation network, while the feature maps obtained from every level of its decoder arm were passed to the discriminator network which acted as a domain classifier. This was done after either up-sampling or down-sampling the feature maps to match the volume size of the second deepest layer. The final loss function for our second model was therefore made up of the generalised Dice loss and an adversarial loss:

$$\mathcal{L}_{method_2} = \mathcal{L}_{seg} - \alpha \mathcal{L}_{adv_2} \quad (4.2)$$

where α is a hyperparameter increased linearly from 0 to 0.05 starting at epoch 20, and which remains equal to 0.05 from epoch 50 onward. In equation 4.2, \mathcal{L}_{adv_2} is the domain discriminator’s classification loss defined as the CE loss (equation 3.14 from Section 3.2.2) between predicted and assigned target labels representing the two domains:

$$\begin{aligned} \mathcal{L}_{adv_2} = & -\log(D(\text{get_feature_maps}(\text{Seg}(x^T)))) \\ & -\log(1 - D(\text{get_feature_maps}(\text{Seg}(x^S)))) \end{aligned} \quad (4.3)$$

Here, `get_feature_maps(Seg(.))` retrieves the feature maps of the decoder arm of the segmentation network `Seg` after either having a target sample as input (x^T), or a source sample as input (x^S).

Similar to Kamnitsas *et al.* [330] we looked at the behaviour of our discriminator and segmentation network when training with different values of $\alpha \in [0.02, 0.05, 0.1, 0.2, 0.5]$. We found the discriminator’s accuracy during training stable for all investigated values, while the segmentation network achieved the lowest loss when $\alpha = 0.05$. The segmentation network was trained similarly to the baseline model, while the discriminator network was trained using the Adam optimiser with $\beta_1 = 0.5$ and $\beta_2 = 0.999$, and a linearly decaying learning rate scheduler starting from $2 \cdot 10^{-3}$.

The generator network used in the image space domain adaptation approach was trained to produce synthesised ePrime volumes, while the segmentation network was trained using the same loss function, optimizer and learning rate scheduler as in the other two methods. In the previous model (adversarial domain adaptation in the latent space) we fed both dHCP and ePrime volumes to the segmentation network to obtain data agnostic feature maps. For this reason, and to allow for a fair comparison between the two unsupervised domain adaptation models, we trained the segmentation network from the image space model on both real dHCP and synthesised ePrime volumes. For both the discriminator and the generator networks the Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ was used, together

with a linearly decaying learning rate scheduler starting from $2 \cdot 10^{-3}$. The loss function of the discriminator was similar to that of the Least Squares GAN [361]: $\mathcal{L}_D = \mathbb{E}_{x \sim T}[(D(x) - b)^2] + \mathbb{E}_{x \sim S}[(D(G(x)) - a)^2]$ where a signified the label for synthesised volumes and b was the label for real volumes. The generator and the segmentation network were trained together using the following loss:

$$\mathcal{L}_{method_3} = \mathcal{L}_{seg} + \mathcal{L}_{adv_3} + \mathcal{L}_{NCC}(G(x), x) \quad (4.4)$$

where $\mathcal{L}_{adv_3} = \mathbb{E}_{x \sim S}[(D(G(x)) - b)^2]$. The additional NCC loss was used between the real and the generated volumes in order to constrain the generator to produce realistic looking ePrime-like images. Without the additional NCC loss, the generator tends to produce images with an enlarged CSF boundary in order to match the preterm-only distribution found in the ePrime dataset, as we have previously shown in [358].

These three methods were trained with and without data augmentation for 100 epochs, during which we used the validation sets to inform us about our models' performance and to decide on the best performing models. For data augmentation we applied: random affine transformations (with rotation angles $\theta_i \sim \mathcal{U}(-10^\circ, 10^\circ)$ and/or scaling values $s_i \sim \mathcal{U}(0.8, 1.2)$), random motion artefacts (corresponding to rotations of $\theta_i \sim \mathcal{U}(-2^\circ, 2^\circ)$ and translations of $t_i \sim \mathcal{U}(-2 \text{ mm}, 2 \text{ mm})$), and random MRI spike and bias field artifacts [206]. The cortical parcellation network was trained in a similar fashion as the baseline tissue segmentation network, with data augmentation in the form of random affine transformations (with the same parameters as above).

4.3 Results

We use the test set to report our final models' results and to also investigate their capability to generalize on the source domain. Finally, we produce tissue segmentation maps for all the subjects in our datasets, and use them as input into ANT's DiReCT algorithm [362] to compute cortical thickness measures. To validate our results, we compare cortical thickness measures between subsets of the two cohorts matched for GA and PMA, for which we expect no significant difference in cortical thickness if the harmonisation was successful. We also assess the association between PMA and cortical thickness in the two cohorts.

4.3.1 dHCP test dataset

Baseline and domain adaptation models. In our first experiment we looked at the performance of the six trained models when applied to data from the source (dHCP) test dataset. The aim was to assess whether our trained models were able to generalise to unseen data from the source domain (dHCP) for which we have

reliable Draw-EM outputs. Figure 4.5 summarizes the results of our trained models, showing mean Dice scores, mean Hausdorff distance calculated using SimpleITK [363, 364], precision and recall. These metrics were computed between the predicted tissue segmentation maps and the Draw-EM output labels for each of the six trained models. The model that obtained the best score is highlighted with the yellow diamond for each metric and tissue type. In terms of Dice scores, out of the six models, the *baseline with augmentation* and *image with augmentation* methods performed best on the source domain test dataset for CSF, dGM, cerebellum and brainstem, with no significant difference between them. For cGM and WM, the best performance was obtained by the *baseline with augmentation* model, while the domain adaptation methods showed a slight decrease in performance. The three models trained without augmentation always performed significantly worse than their augmented counterparts. In terms of average Hausdorff distance, both the *baseline with*

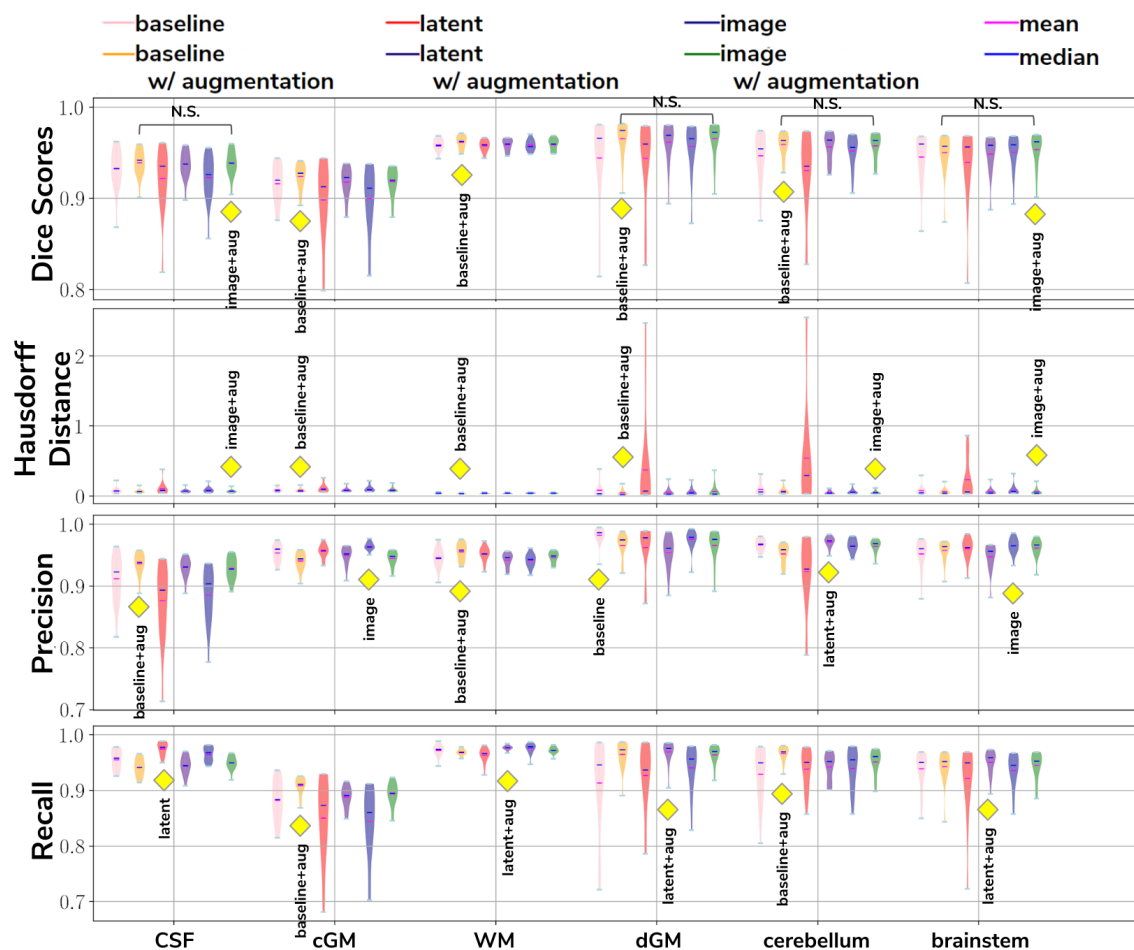


Figure 4.5: The results on our dHCP test dataset for all six methods. The yellow diamond highlights the model which obtained the best mean score for its respective tissue type and metric. Models which obtained non-significant differences when compared to the best performing method are shown above each pair.

augmentation and *image with augmentation* models performed well, while the *latent without augmentation* model performed worse than all the other models for all tissue types. Highest precision scores were obtained by the *baseline with augmentation* model for both CSF and WM, the *image without augmentation* method for both

cGM and brainstem, the *baseline without augmentation* for dGM, and the *latent with augmentation* model for cerebellum. Highest recall scores were obtained by the *baseline with augmentation* model for cGM and cerebellum, the *latent with augmentation* model for WM, dGM and brainstem, and the *latent without augmentation* model for CSF. These results show that our trained models were able to generalise to unseen data from the source domain, and that the performance on the dHCP dataset was not compromised by using domain adaption techniques.

Cortical parcellation network. To assess the performance of our trained cortical parcellation network, we applied it on the source (dHCP) test dataset, where the inputs were binary Draw-EM cortical gray matter tissue maps. For each subject in our test dataset, the network produced a 12-channel output, consisting of: frontal left, frontal right, cingulate, temporal left, temporal right, insula left, insula right, parietal left, parietal right, occipital left, occipital right, and background, respectively. Table 4.2 summarizes these results in terms of minimum, maximum and mean Dice scores for each of the 11 cortical substructures. When compared with the Draw-EM outputs [148], the network obtained an overall mean Dice score of 0.97.

Tissue	min	max	mean	Tissue	min	max	mean
Frontal (left)	0.98	0.99	0.99	Frontal (right)	0.98	0.99	0.99
Temporal (left)	0.96	0.99	0.98	Temporal (right)	0.97	0.98	0.98
Insula (left)	0.95	0.97	0.96	Insula (right)	0.95	0.97	0.96
Parietal (left)	0.96	0.98	0.97	Parietal (right)	0.96	0.98	0.97
Occipital (left)	0.94	0.98	0.97	Occipital (right)	0.95	0.98	0.97
Cingulate	0.93	0.97	0.96				

Table 4.2: Dice Scores obtained on the dHCP test set for the trained cortical parcellation network.

4.3.2 Validation of data harmonisation

In order to evaluate the extent to which each of the trained models managed to harmonise the segmentation maps of the two cohorts, we looked at tissue volumes and mean cortical thickness measures between subsamples of the dHCP ($N = 30$; median GA = 30.50 weeks; median PMA = 41.29 weeks) and ePrime ($N = 30$; median GA = 30.64 weeks; median PMA = 41.29 weeks) cohort which showed comparable GA at birth and PMA at time of scan (see Table 4.1). A direct comparison between the two cohort subsets shows that the dHCP and ePrime neonates did not differ significantly in terms of sex ($\chi^2(1) < 0.001$, $p > 0.05$), or maternal ethnicity ($\chi^2(4) = 4.32$, $p > 0.05$), coded as “white or white British”, “black or black British”, “asian or asian British”, “mixed race”, and “other”. As a proxy for socio-economic status, we derived an Index of Multiple Deprivation (IMD) score based on parental postcode at the time of infant birth (Department for Communities and Local Government, 2011³). This measure is based on seven domains of deprivation within

³<https://tools.npeu.ox.ac.uk/imd/>

each neighbourhood compared to all others in the country: income, employment, education, skills and training, health and disability, barriers to housing and services, living environment and crime. Higher IMD values therefore indicate higher deprivation. IMD score did not differ significantly between dHCP ($M = 21.4$, $SD = 10.7$) and ePrime ($M = 18.0$, $SD = 11.6$) subsets, suggesting that these two groups are comparable in terms of environmental background.

For these two cohort subsamples with similar GA and PMA, we expected both volumes and cortical thickness measures not to differ after applying the harmonisation procedures. We also investigated the relationship between PMA and volumes and cortical thickness respectively, before and after applying the harmonisation. Linear regressions were performed in the comparable data subsets testing the effects of PMA and cohort on volumes (or cortical thickness), controlling for GA and sex.

Volumes. Figure 4.6 shows the tissue volumes for both the original and the predicted segmentations. Significant volume differences between the two subsamples (i.e., significant effect of cohort in the regression model) are reported above each tested model. To summarise, the *image with augmentation* model performed best, by showing no significant differences in the two cohorts for cortical gray matter, white matter, deep gray matter, cerebellum and brainstem. The cerebrospinal fluid volumes were significantly different between the two cohorts for all our trained models, as well as for the original ePrime segmentation masks.

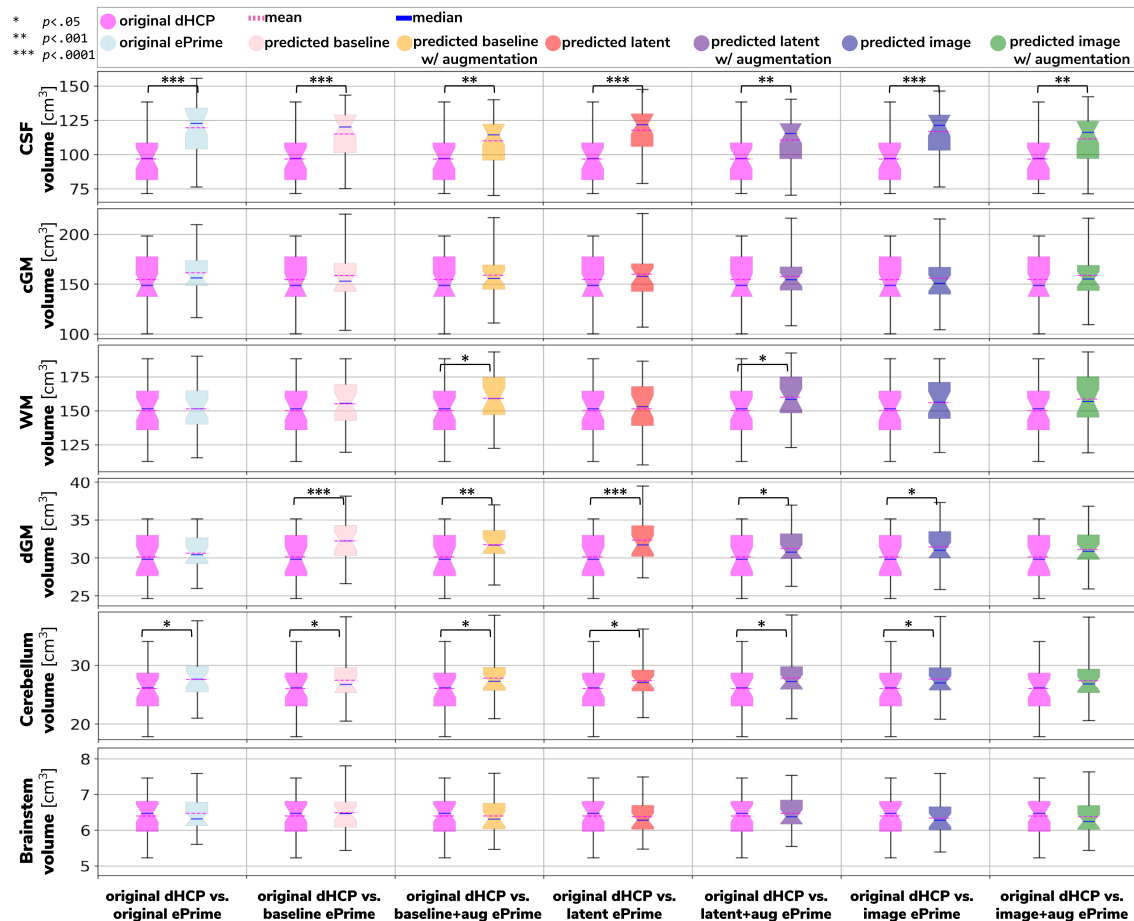


Figure 4.6: Comparison of volume measures for our 6 tissue types (CSF, cGM, WM, dGM, cerebellum and brainstem) between original Draw-EM dHCP segmentations and original Draw-EM ePrime segmentations (first column), or between original Draw-EM dHCP segmentations and ePrime segmentations obtained with the 6 trained models (columns 2 - 7). Linear regressions were performed in the comparable data subsets testing the effects of cohort on volumes, controlling for PMA, GA, and sex (volume \sim cohort + PMA + GA + sex). The asterisks indicate a statistically significant effect of cohort in the linear regression.

Cortical thickness. Figure 4.7 summarizes the results of applying the cortical thickness algorithm on the predicted segmentation maps for all six methods. Before harmonisation, the matched subsets from the dHCP and ePrime cohorts showed a significant difference in mean cortical thickness (dHCP: $M = 1.73$, $SD = 0.12$; ePrime: $M = 1.93$, $SD = 0.13$; $t(58) = 6.33$, $p < .001$). After applying the harmonisation to the ePrime sample, mean cortical thickness no longer differed between the two subsamples for four of our methods. These results are summarised in panel H from Figure 4.7, where the models which obtained harmonised values in terms of mean cortical thickness measures are shown in bold.

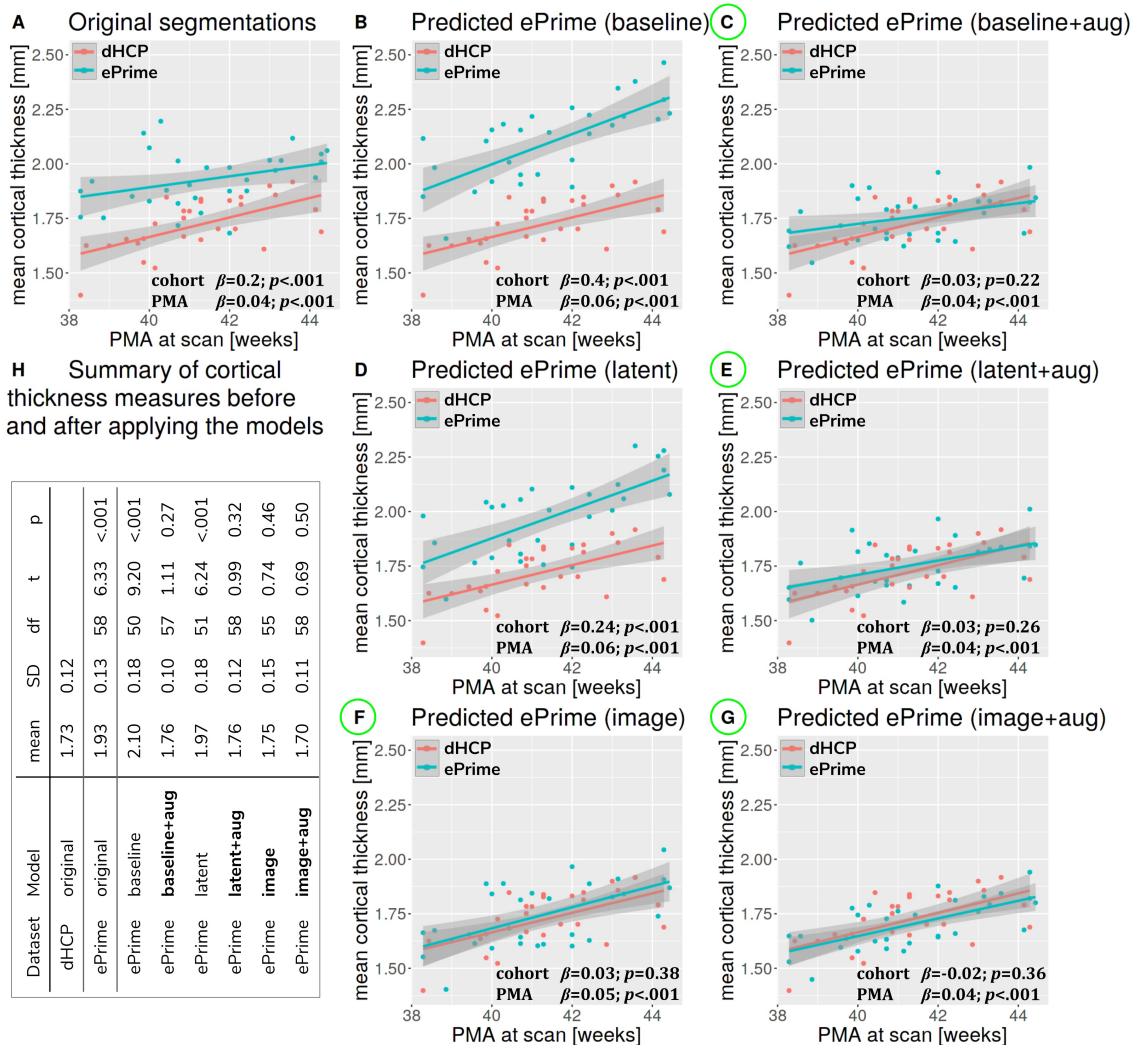


Figure 4.7: The association between PMA and mean cortical thickness before (A) and after (B-G) applying the data harmonisation models on the matched dHCP and ePrime subsets. A linear model regressing mean cortical thickness measures on PMA, GA, sex, and cohort revealed a significant effect of cohort for the original segmentations (A), and the predicted maps (B - *baseline without augmentation* and D - *latent without augmentation*). The effect of cohort was rendered non-significant for four of the methods (C - *baseline with augmentation*, E - *latent with augmentation*, F - *image without augmentation* and G - *image with augmentation*). Panel H summarizes cortical thickness measures before and after applying the models.

Figure 4.7 also shows the association between PMA and mean cortical thickness before (panel A) and after applying the models (panels B-G) on the matched dHCP and ePrime subsets. A linear model regressing unharmonised mean cortical thickness on PMA, GA, sex, and cohort revealed a significant effect of cohort ($\beta = 0.20$; $p < .001$), consistent with a group difference in mean cortical thickness reported above, as well as a significant effect of PMA ($\beta = 0.04$; $p < .001$), consistent with an increase in cortical thickness with increasing PMA. After applying the methods, the effect of cohort was rendered non-significant for four of the methods (see highlighted panels C, E, F, G from Figure 4.7), while the effect of PMA remained stable across all six methods.

We performed a similar analysis on thickness measures of the cortical substructures. To obtain these measures, we used the original and the predicted cortical gray matter segmentation maps (obtained by applying each of our six methods) as input to the trained cortical parcellation network to predict cortical substructure masks. We then used these masks to calculate local cortical thickness measures. Our results are summarised in Figure 4.8.

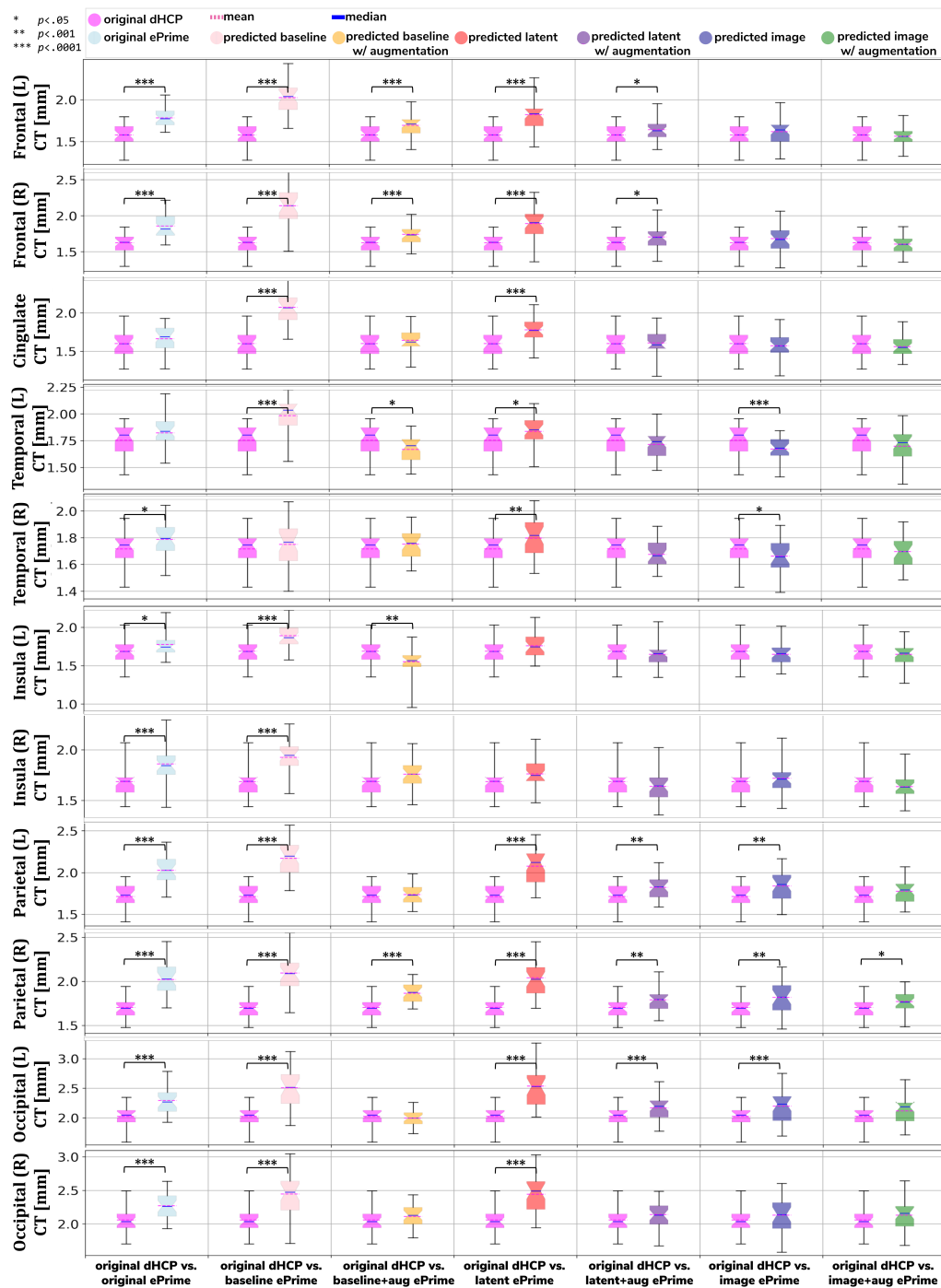


Figure 4.8: Comparison of local mean cortical thickness measures between original Draw-EM dHCP segmentations and original Draw-EM ePrime segmentations (first column), or between original Draw-EM dHCP segmentations and ePrime segmentations obtained with the 6 trained models (columns 2 - 7). Linear regressions were performed in the comparable data subsets testing the effects of cohort on local cortical thickness measures, controlling for PMA, GA, and sex ($CT \sim \text{cohort} + \text{PMA} + \text{GA} + \text{sex}$). The asterisks indicate a statistically significant effect of cohort in the linear regression.

Qualitative assessment of predicted segmentation maps. To further narrow down which of the four remaining methods was best at harmonising our ePrime neonatal dataset, we looked at the predicted segmentations. Figure 4.9 shows two example neonates from the ePrime dataset with GA = 32.9w, PMA = 43.6w, and with GA = 28.7w, PMA = 44.7w, respectively. The first column shows T_2w sagittal and axial slices, respectively, while the following four columns show example tissue prediction maps produced by the four models: *baseline with augmentation*, *latent with augmentation*, *image* and *image with augmentation*, respectively. On the first row we show an example neonate for which three of the models (*baseline with augmentation*, *latent with augmentation* and *image*) misclassified a part of the cortex as being deep gray matter. This is more pronounced in the *baseline with augmentation* model, while the *latent with augmentation* and *image* show a slight improvement. The *image with augmentation* model corrected the problem entirely. On the second row the yellow arrow points to an area of CSF where the *baseline with augmentation* model misclassified it as dGM, while the other three models did not have this problem. The red arrow on the other hand points to an area where the *latent with augmentation* model misclassified cGM as deep gray matter. This problem does not appear in the other models.

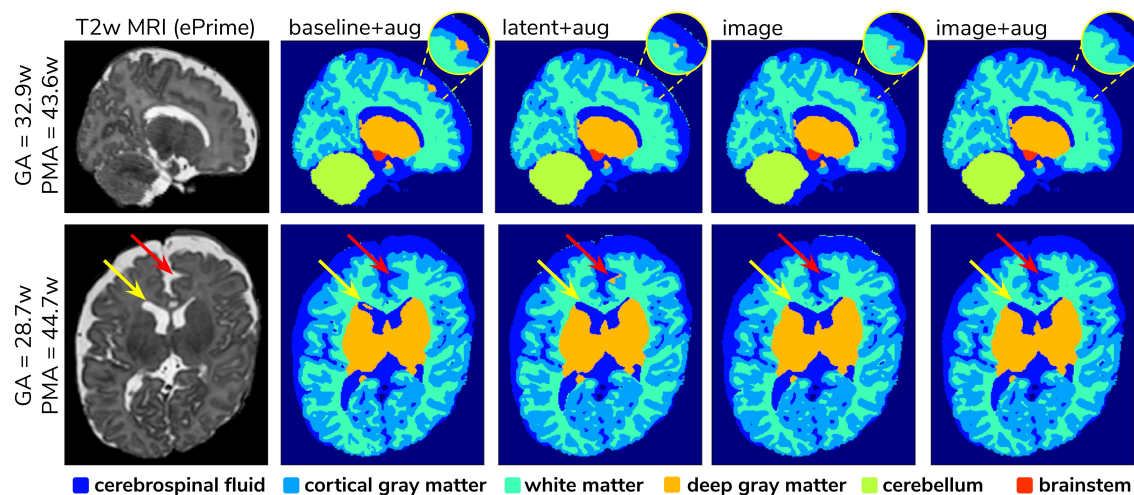


Figure 4.9: Example predicted segmentation maps for the best performing models.

Figure 4.10 shows the axial, sagittal and coronal slices of an ePrime neonate (GA = 32.86w and PMA = 39.86w). The first line shows the T_2w MR image, while the second and third lines show the CSF boundary of both the Draw-EM algorithm and the *image with augmentation* method. The green arrows point to a WM region where the Draw-EM algorithm performed worse (classified the area as CSF) than our proposed model. This problem was corrected by the *image with augmentation* method.

Although all four methods performed well in terms of harmonising tissue segmentation volumes and global mean cortical thickness values for the two subsamples with similar GA and PMA, previously presented quantitative results as well as the example above suggest that the *image with augmentation* method was more robust.

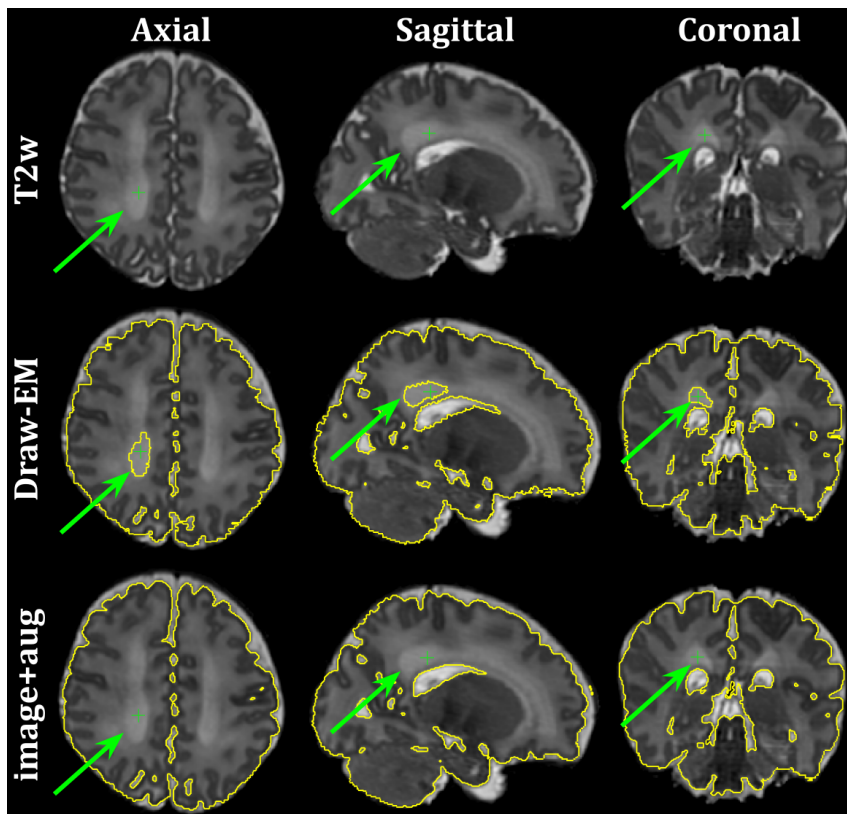


Figure 4.10: Example of a neonate from the ePrime dataset with 32.86 weeks GA at birth and 39.86 weeks PMA at scan. The green arrow points to a region which was segmented as CSF by Draw-EM, but then corrected by our model.

4.3.3 Analysis of harmonised cortical substructures

In this section we analyze the harmonised cortical gray matter segmentation maps using the *image with augmentation* model. We produce tissue segmentation maps for the entire ePrime dataset and calculate cortical thickness measures on the predicted and Draw-EM cortical gray matter tissue maps of both cohorts. In addition, we use the trained cortical parcellation network to produce cortical substructure masks. We perform a term *vs* preterm analysis on the harmonised cortical gray matter maps and we show the importance of harmonising the data with a proof-of-principle application setting where we investigate the association between cortical thickness and a language outcome measure.

Comparison of term and preterm cortical maps. Associations between cortical thickness and GA or PMA in the full dHCP and ePrime datasets (excluding subjects with PMA > 45 weeks) for the whole cortex are depicted in Figure 4.11, where we show individual regression lines for preterm-born and term-born neonates. The first column consists of dHCP-only subjects, while the following two columns showcase both cohorts together, before and after harmonising the cortical gray matter tissue maps.

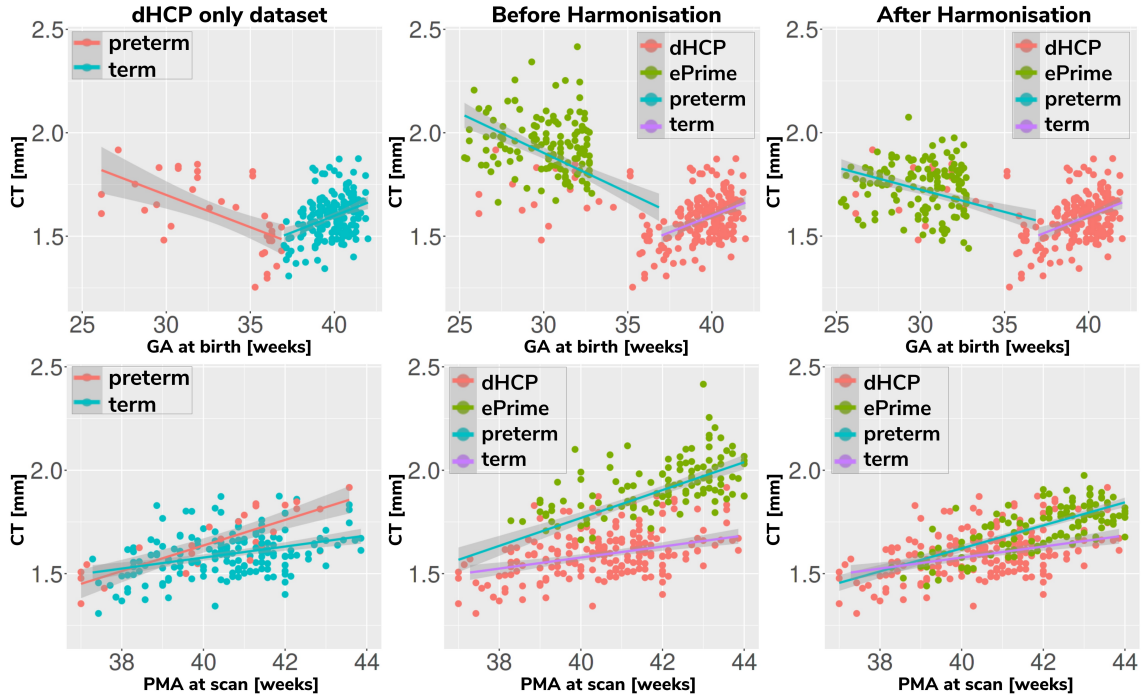


Figure 4.11: Mean cortical thickness measures in our dHCP dataset (first column), and in both cohorts before (second column) and after (third column) harmonising the tissue segmentation maps. The first row plots the cortical thickness measures against GA, while the second row plots the cortical thickness measures against PMA, with individual regression lines on top.

A linear model regressing dHCP-only mean cortical thickness on PMA, GA, sex, birth weight and the interaction between PMA and GA revealed a significant effect of PMA ($\beta = 0.19$; $p < 0.001$), a significant effect of GA ($\beta = 0.16$; $p = 0.002$), and a significant effect of the interaction between PMA and GA ($\beta = -0.004$; $p = 0.002$), indicating that infants born at a lower GA showed a stronger relationship between PMA and CT. When performing the same analysis in the pooled ePrime and dHCP data before harmonising the maps, the effect of GA and the effect of the interaction were rendered not significant (GA: $\beta = 0.009$; $p = 0.7$ and PMA*GA: $\beta = -0.0006$; $p = 0.5$, respectively). This is corrected after harmonising the tissue maps, where the effects of GA ($\beta = 0.06$; $p = 0.02$) and the effects of the GA and PMA interaction ($\beta = -0.001$; $p = 0.02$) are, again, significant.

The second and third columns of Figure 4.11 show that after harmonising the tissue segmentation maps, the ePrime preterm-born neonates (green dots) are brought downwards into a comparable range of values to the dHCP preterms (red dots). Moreover, when plotting the cortical thickness measures against PMA, after harmonising the tissue maps, the intersection between the two individual regression lines (term and preterm-born neonates) happens at roughly the same age (PMA = 38.5 weeks) as in the dHCP-only dataset.

We extended the term *vs* preterm analysis on cortical thickness substructures. Figure 4.12 shows the results of applying a linear model regressing mean cortical

thickness measures on PMA, GA, sex, birth weight and prematurity, where significant differences ($p < 0.05$) between the two cohorts (term and preterm-born neonates) are highlighted in the image.

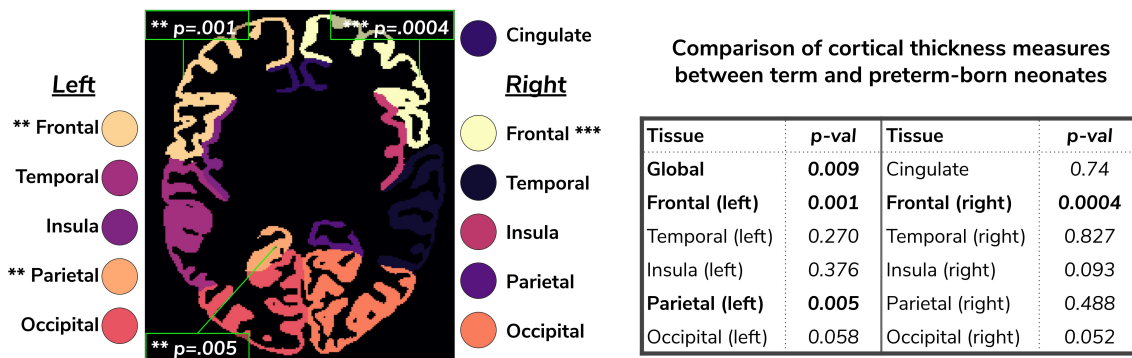


Figure 4.12: Comparison of cortical thickness measures for the whole cortex and for each of the 11 cortical subregions between term and preterm-born neonates. The results of the linear regression are reported in the table in terms of differences between term and preterm-born neonates.

Behavioural outcome association. As a final proof-of-principle, we demonstrate the importance of data harmonisation in an application setting investigating the association between neonatal cortical thickness and a behavioural outcome measure (see Figure 4.13). For this, we consider language abilities as assessed between 18 and 24 months in both dHCP and ePrime cohorts using the Bayley Scales of Infant and Toddler Development [365]. Age-normed composite language scores were available for 203 toddlers from the dHCP cohort ($M = 96.43$; $SD = 14.89$) and 136 toddlers from the ePrime cohort ($M = 91.25$; $SD = 17.37$). For the neonatal cortical thickness measure, we focus on the left and right frontal cortex for illustration.

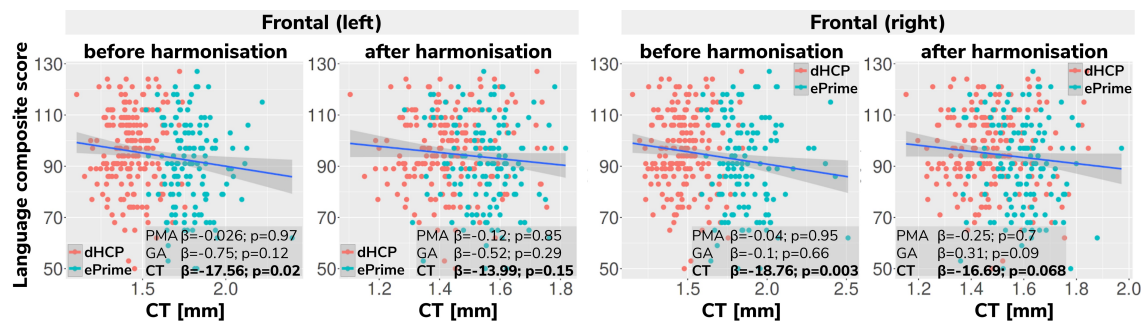


Figure 4.13: Language composite score against predicted left and right frontal cortical thickness measures before and after harmonising the tissue segmentation maps. Without harmonisation (columns 1 and 3) there appears to be a significant association between left or right frontal cortical thickness and language abilities, but after harmonisation (columns 2 and 4) the effect of cortical thickness on language ability is rendered non-significant in both left and right frontal cortex. This demonstrates the importance of data harmonisation without which pooling images from separate datasets can lead to spurious findings that are driven by differences in acquisitions rather than by true underlying effects.

Regressing composite language score against left or right frontal cortical thickness in each cohort separately, controlling for PMA, GA, sex and intracranial volume

showed that there was no significant association between neonatal left/right frontal cortical thickness and language abilities at toddler age in either of the cohorts. However, when pooling data from both cohorts together and rerunning the same analysis (using un-harmonised cortical thickness measures), a significant association between left/right frontal cortical thickness and language abilities is seen (left: $\beta = -17.56$, $p < 0.05$, right: $\beta = -18.76$, $p < 0.05$), suggesting that greater frontal cortical thickness at term-equivalent age is associated with reduced language abilities at toddler age.

However, as can be seen in Figure 4.13, this is likely a spurious effect due to (artefactually) heightened cortical thickness values in un-harmonised ePrime data combined with lower language composite scores in the ePrime cohort (consistent with effects typically observed in preterm cohorts). Indeed, when rerunning the same analysis on harmonised data pooled across both cohorts, the effect of cortical thickness on language ability is rendered non-significant in both left ($\beta = -13.99$, $p = 0.15$) and right ($\beta = -16.69$, $p = 0.068$) frontal cortex, consistent with the ground-truth findings in each individual cohort.

4.4 Discussion and future work

In this work we studied the application and viability of unsupervised domain adaptation methods for harmonising tissue segmentation maps of two neonatal datasets (dHCP and ePrime). Our aim was to obtain volumetric and cortical thickness measures that are only affected by brain anatomy and not by the acquisition protocol or scanner, in order to improve the statistical power of imaging or imaging-genetic studies. We proposed an image-based domain adaptation model where a tissue segmentation network was trained with real dHCP and synthesised ePrime T_2w 3D MRI volumes. The generator network was trained to produce realistic images in order to fool a domain discriminator, while also minimizing an NCC loss which aimed to enforce image similarity between real and synthesised images [358]. We trained this model using dHCP Draw-EM segmentation maps, and we compared it with a baseline 3D U-Net [215], and a latent space domain adaptation method [330]. The three methods were trained with and without data augmentation [206].

First, we looked at the performance of each of the six trained models on the source (dHCP) test dataset, by comparing predicted tissue segmentation maps with the Draw-EM output labels, with the aim of measuring fidelity of our trained segmentation methods for the original dHCP domain. Our results on the source (dHCP) test dataset suggest that our trained models were able to generalise to unseen data from the source domain. At the same time, Dice score results on the test set for the proposed *image with augmentation* model are high and are similar in performance when compared with the *baseline with augmentation* method. This suggests that adding the contrast transfer step does not diminish the quality of the segmentations.

We then analysed the extent to which each of the 6 trained models managed to harmonise the tissue segmentation maps of our two cohorts, by looking at tissue volumes and mean cortical thickness measures between subsamples of the dHCP and ePrime cohorts which showed comparable GA at birth and PMA at time of scan, as well as similar gender and maternal ethnicity. Our results showed that our proposed model (*image with augmentation*) harmonised the predicted tissue segmentation maps in terms of cortical gray matter, white matter, deep gray matter, cerebellum and brainstem volumes (Figure 4.6). In terms of mean global cortical thickness measures, four of the trained methods (*baseline with augmentation*, *latent with augmentation*, *image* and *image with augmentation*) achieved comparable values when compared to the dHCP subset. In fact, we hypothesize that these four methods provided the best overall results because either they were trained using data augmentation or they acted as a deep learning-based augmentation technique [366], which made the segmentation network more robust to the different contrast, population bias and acquisition protocol of the ePrime dataset.

Using the cortical parcellation network, we also produced cortical thickness measures for the 11 cortical subregions (see Tables 1 and 2). Again, the models trained with augmentation performed better than their no augmentation counterparts (see Figure 4.8). However, our proposed *image with augmentation* model performed best, whereby ePrime values, tending towards higher values before harmonisation, were brought downwards into a comparable range of values to dHCP, for 10 out of 11 cortical subregions (see Figure 4.8 last column). For the right parietal lobe, our proposed method outperformed the original segmentations and the other 5 models, but did not manage to bring the values down to a non-significant range. One potential reason for this is that, on a visual inspection, the ePrime cohort appears to suffer from more partial volume artifacts than its dHCP counterpart, which can confuse the segmentation network and can lead to overestimation of the cortical gray matter / cerebrospinal fluid boundary.

A close inspection of the predicted tissue segmentation maps (see Figure 4.9) also showed that our proposed model (*image with augmentation*) corrected misclassified voxels which were prevalent in the other 3 methods. At the same time, the proposed *image with augmentation* method outperformed the original DrawEM segmentation by correcting a region of WM which was wrongly classified as CSF (see Figure 4.10). Our results suggest that, in terms of consistency of volumes and regional cortical thickness measures derived from dHCP and ePrime neonates (Figure 4.6 and Figure 4.8), as well as the qualitative examples (Figure 4.9 and Figure 4.10), our proposed *image with augmentation* model resulted in more consistent outputs than the other methods.

We used the harmonised cortical segmentation maps to look at differences in both global and local cortical thickness measures between term and preterm-born neonates. We showed in Figure 4.12 that our harmonised cortical gray matter maps resulted in global thickness measures which were comparable with the dHCP-only neonates, while also revealing a significant effect of GA and the interaction between age at scan and at birth. We performed a similar analysis on the local cortical

thickness measures and highlighted three regions of interest (frontal left, frontal right, and parietal left) which showed significant differences between the two cohorts (see Figure 4.12). These regions are consistent with previous studies [367] where cortical thickness measures were shown to differ in preterm-born neonates when compared to term-born neonates in an adolescent cohort.

Finally, we showed the importance of harmonising the cortical tissue maps by investigating the association between neonatal cortical thickness and a language outcome measure. After harmonisation, regressing language composite score against predicted left or right frontal cortical thickness in the two pooled datasets, showed no significant effect of cortical thickness (second column of Figure 4.13), consistent with the ground-truth results seen in each cohort individually. This analysis demonstrates that without data harmonisation, pooling images from separate datasets can lead to spurious findings that are driven by systematic differences in acquisitions rather than by true underlying effects. Our harmonisation allows for our two datasets to be combined into joint analyses while preserving the underlying structure of associations with real-world outcomes.

Recently, it has become increasingly commonplace to share imaging data amongst research communities to form large and diverse datasets [368]. As a result of this collective endeavour, data harmonisation becomes a critical component for enabling the development of novel biomarkers that are invariant to different imaging equipment or patient demographics. This is also an essential step towards translating neuroimaging research into the clinics. Our model, however, lacks a number of important criteria to be fully deployed into clinical practice. First, the data used in this project was preprocessed (alignment to a common template, skull-stripping using already available Draw-EM brain masks, and linear upsampling to a 0.5 mm isotropic resolution), which means that in a clinical setting it would need to be part of a more complex framework, in order to become a fully automated pipeline, and be robust to the variability present in neonatal brain MRI. Second, validating our model’s performance is crucial before deploying it in a clinical environment. More specifically, ground truth labels obtained with the help of clinical experts would be required in order to assess whether the harmonized images are able to accurately represent the underlying anatomical structures and tissue properties. Finally, deploying our proposed data harmonization model in a clinical setting requires further training with a larger cohort of representative neonatal MR images, of different scanner manufacturers, acquisition parameters, imaging protocols, and patient demographics.

Our study suffers from several limitations. First, one particular issue is our reliance on accurate dHCP labels. More specifically, we assume that the dHCP segmentation maps which were obtained through an automated process (Draw-EM), and therefore the computed cortical thickness measures, are the *baseline standard* that the predicted ePrime labels use as reference. Moreover, the ePrime dataset had no available tissue maps obtained through other means, such as manual segmentations, as the same automated method (Draw-EM) did not yield good results when applied to ePrime (see example in Figure 4.10). Second, this study was focused on

single-source unsupervised domain adaptation approaches, which might limit application in terms of applying the method to a different neonatal dataset. However, by utilising reliable tissue segmentation maps from multiple neonatal databases, the proposed model can be extended to a multi-source domain adaptation pipeline [369, 370].

Moreover, recent literature [371, 338] suggests that unsupervised models can reach undesirable results when the two domains are highly disparate, and propose the use of a small amount of labelled target domain data during training. Also, Zhang *et al.* [372] demonstrated that semi-supervised learning can outperform unsupervised domain adaptation models on classification benchmarks. In future, we would like to investigate semi-supervised approaches by including reliable segmentations of the ePrime cohort, and evaluate their performance when compared to our proposed unsupervised DA model. Additionally, the latent based domain adaptation method was trained using the features at every layer of the decoding branch, without analysing different combinations of the encoding-decoding layers. Future work will therefore aim to systematically evaluate our design choices via ablation studies. Finally, we focused our work on investigating structural (T_2w) datasets only, and in future we aim to extend this study to harmonise diffusion data as well.

Attention-driven multi-channel deformable registration of structural and microstructural neonatal data

Motivation

Accurate alignment of neonatal MRI in presence of rapid development is needed for downstream tasks.

Contribution

A novel multi-channel deep learning image registration framework that aims to combine information from T_2w neonatal scans with DWI-derived FA maps.

Publications

- Grigorescu, I. et al. (2021). *Uncertainty-Aware Deep Learning Based Deformable Registration*. UNSURE 2021. LNCS (Springer)
[🔗 doi.org/10.1007/978-3-030-87735-4_6](https://doi.org/10.1007/978-3-030-87735-4_6)
- Grigorescu, I. et al. (2022). *Attention-Driven Multi-channel Deformable Registration of Structural and Microstructural Neonatal Data*. PIPPI 2022. LNCS (Springer)
[🔗 doi.org/10.1007/978-3-031-17117-8_7](https://doi.org/10.1007/978-3-031-17117-8_7)

Code available at:

[🔗 github.com/irinagrigorescu/attentionneonatalmrireistration](https://github.com/irinagrigorescu/attentionneonatalmrireistration)

5.1 Introduction

The neonatal brain undergoes dramatic changes during early life, such as cortical folding and myelination. Non-invasive MRI offers 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 [60] or diffusion [61]). Studies have shown that combining diffusion and structural data to drive the registration [54, 57, 56, 55] improves the overall alignment. Classic approaches for fusing these channels are based on simple averaging of the deformation fields from the individual channels [54], or weighting the deformation fields based on certainty maps calculated from normalised gradients correlated to structural content [57, 56, 51].

In order to establish accurate correspondences between MR images acquired during the neonatal period, I propose an attention-driven multi-channel deep learning image registration framework that aims to combine information from T_2w neonatal scans with DWI-derived FA maps. The proposed solution is based on a diffeomorphic framework for non-rigid registration through stationary velocity field representation [252]. This is of interest when registering images of the developing brain [357] as diffeomorphisms lead to one-to-one transformations which are topology preserving and inverse consistent [69]. Moreover, these properties are important for the construction of brain atlases [373], as well as enabling plausible downstream analysis of brain volume, shape, and change over time [69]. On top of this, the proposed model also takes advantage of learnt spatial attention maps to select the most salient features from both T_2w and FA maps.

More specifically, a CVAE image registration network is trained to align either structural or microstructural data to a 36 weeks neonatal atlas [51] of the same modality. As a second step, a CNN, which learns attention maps for weighted combination of the predicted modality-specific velocity fields, is trained to achieve an optimal multi-channel alignment. Throughout this work, 3D MRI brain scans [11] acquired as part of the dHCP¹ are used as the moving images, while 36 weeks neonatal multi-modal atlas² [51] is used as the fixed image.

To evaluate the proposed framework, a test set of 30 neonates scanned around 40 weeks PMA is used. Moreover, a comparison study is performed to evaluate the results against registration networks trained on T_2w -only, FA-only, and both modalities at the same time, either with or without attention. We also explored the use of visual attention network blocks [309, 314] and our previously proposed uncertainty-aware mechanism [374]. The quantitative evaluation confirmed that while cortical structures were better aligned using T_2w data and white matter tracts

¹developingconnectome.org

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

were better aligned using FA maps, the attention-based multi-channel registration aligned both types of structures accurately.

5.2 Methods

5.2.1 Data acquisition and preprocessing

The MRI data used in this study was collected as part of the dHCP project [11] and details about the data acquisition can be found in Section 1.2.3. In total, we use 414 3D T_2w volumes and FA maps of neonates born between 23 – 42 weeks GA and scanned at term-equivalent age (37–45 weeks PMA). As preprocessing steps, we first affinely pre-registered the data to a common 36 weeks gestational age atlas space [51] using the MIRTk software toolbox [68], and then we resampled both structural and microstructural volumes to be 1 mm^3 isotropic resolution. To obtain the FA maps, we used the MRtrix3 toolbox [375], and we performed skull-stripping using the available dHCP brain masks [148]. Finally, we cropped the resulting images to a $128 \times 128 \times 128$ size.

To train our models, we first performed an 80 – 10 – 10% split of our dataset, resulting in 350 subjects for training, 34 for validation and the remaining 30 subjects for test, as described in Table 5.1. This was achieved through stratified splitting in order to keep the distribution of ages at scan and the male-to-female ratio close to the original distribution (of the entire dataset of 414 subjects). 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.

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)

Table 5.1: 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)

5.2.2 Network architectures

Multi-channel image registration (*baseline*). In this study, we employ a CVAE [376] to model the registration probabilistically as proposed by [252]. Figure 5.1 shows the network architecture, where, a pair of 3D MRI volumes \mathbf{M}_{T2w} and \mathbf{F}_{T2w}

(or \mathbf{M}_{FA} and \mathbf{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* [182] steps) transform it into a topology-preserving deformation field φ_{T2w} (or φ_{FA}). A *Spatial Transformer* layer [181] is then used to warp (linearly resample) the moving images \mathbf{M}_{T2w} (or \mathbf{M}_{FA}) and obtain the moved image $\mathbf{M}_{T2w}(\varphi_{T2w})$ (or $\mathbf{M}_{FA}(\varphi_{FA})$).

Throughout this work, we use a 36 weeks old neonatal structural (T_2w) and microstructural (FA maps) atlas [51] 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.

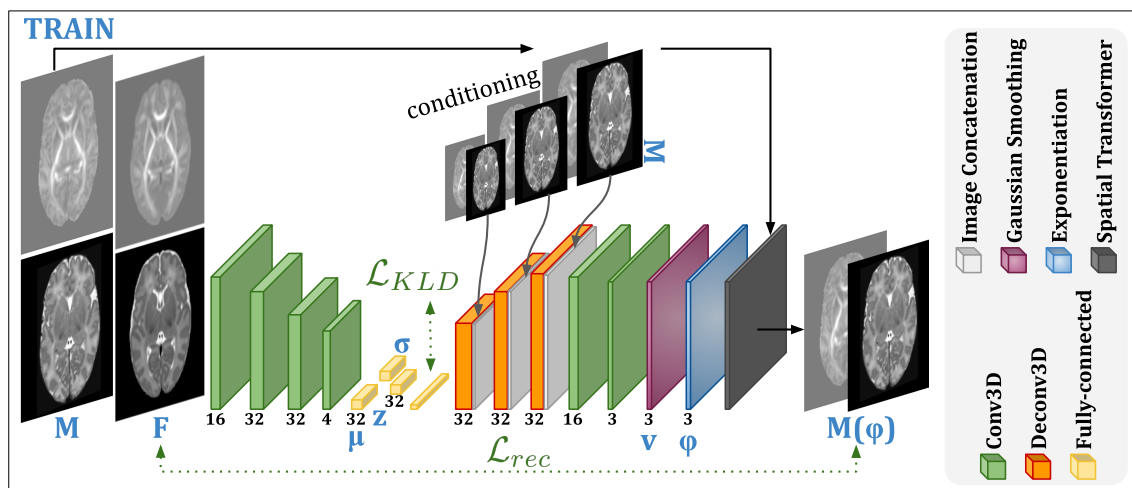


Figure 5.1: Multi-channel image registration network based on the work proposed by Krebs *et al.* [252]. The network takes as input the concatenated moving (\mathbf{M}) and fixed (\mathbf{F}) volumes and produces a velocity field v which is transformed into a deformation field φ through *scaling and squaring* layers. The spatial transformer layer warps the moving image into the moved $\mathbf{M}(\varphi)$. Note that after the concatenation of the moving image(s) in the decoder layers, the number of features becomes $32 + 1$ in the single-channel case, and $32 + 2$ in the multi-channel case.

The network architecture is similar to the original paper [252], but uses a latent code size of 32 and a Gaussian smoothing layer with $\sigma = 1$ mm (kernel size 3^3). More specifically, the encoder branch is made up of four 3D convolutional layers of 16, 32, 32, and 4 filters, respectively, with a kernel size of 3^3 , followed by *Leaky ReLU* ($\alpha = 0.2$) activations [377]. The bottleneck (μ, σ, z) is fully-connected. The decoder branch is composed of three 3D deconvolutional (transpose convolutions) layers of 32 filters and a kernel size of 3^3 each, followed by *Leaky ReLU* ($\alpha = 0.2$) activations. The feature maps of the deconvolutional layers are concatenated with the original-sized or downsampled versions of the moving input volumes. Two more convolutional layers (with 16 and 3 filters, respectively) are added, followed by a Gaussian smoothing layer which outputs the velocity field v .

At inference time, the trained networks are used to generate multiple deformation fields φ_i , as shown in Figure 5.2. More specifically, for each subject in the 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 (here, $\epsilon \sim \mathcal{N}(0,1)$) and pass them through the trained decoder networks to generate n dense deformation fields φ_i . Throughout this work we set $n = 50$. From these, we obtain a mean deformation field $\bar{\varphi}$ and a standard deviation deformation field σ_φ .

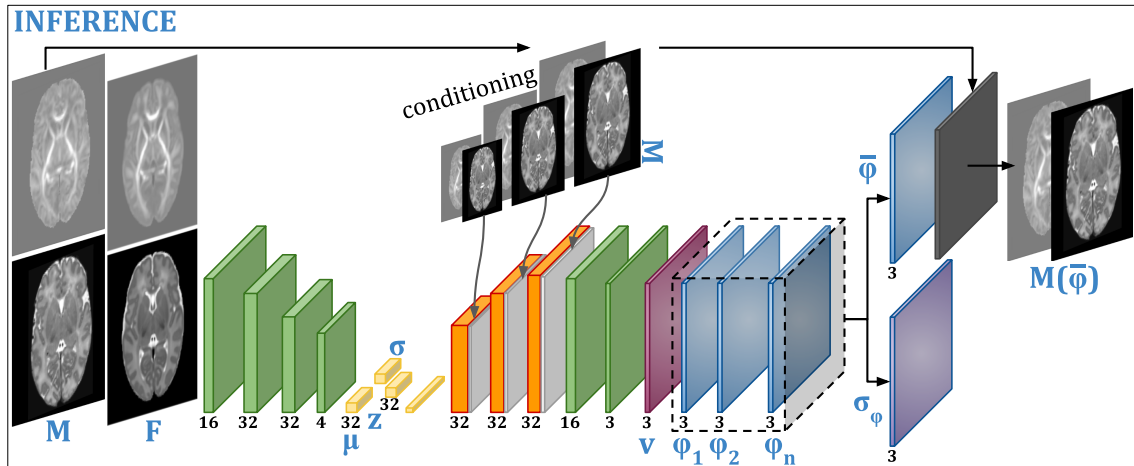


Figure 5.2: At inference time the trained network generates multiple deformation fields φ_i , from which a voxel-wise mean deformation field ($\bar{\varphi}$) and a standard deviation deformation field (σ_φ) can be produced.

Uncertainty-aware registration. To investigate uncertainty-aware image registration, we use our trained models to generate uncertainty maps. This is achieved by combining the pretrained T_2w -only and FA-only models in a three-step process described in Figure 5.3. First, we generate n dense deformation fields φ_i , and create the modality-specific mean deformation fields $\bar{\varphi}_{T_2w}$ and $\bar{\varphi}_{FA}$, and uncertainty maps $\sigma_{\varphi_{T_2w}}$ and $\sigma_{\varphi_{FA}}$. Second, we calculate the certainty maps ($\alpha_{\varphi_{T_2w}}$, $\alpha_{\varphi_{FA}}$) using the following equations:

$$\alpha_{\varphi_{T_2w}} = \frac{1/\sigma_{\varphi_{T_2w}}}{1/\sigma_{\varphi_{T_2w}} + 1/\sigma_{\varphi_{FA}}} ; \alpha_{\varphi_{FA}} = \frac{1/\sigma_{\varphi_{FA}}}{1/\sigma_{\varphi_{T_2w}} + 1/\sigma_{\varphi_{FA}}} \quad (5.1)$$

Note that $\alpha_{\varphi_{T_2w}} + \alpha_{\varphi_{FA}} = \mathbf{1}$ (a tensor of the same size as $\alpha_{\varphi_{T_2w}}$ or $\alpha_{\varphi_{FA}}$, with 1 at every voxel location). Finally, the uncertainty-aware model’s deformation field is constructed by locally weighting the contributions from each modality with the certainty maps:

$$\varphi = \alpha_{\varphi_{T_2w}} \odot \bar{\varphi}_{T_2w} + \alpha_{\varphi_{FA}} \odot \bar{\varphi}_{FA} \quad (5.2)$$

where \odot represents element-wise multiplication.

Attention-driven registration. For the attention-driven registration task, we construct a CNN which uses pairs of modality-specific velocity fields as input, and outputs a combined velocity field which aims to align both structural and microstructural data simultaneously. The network learns the attention maps α_{T_2w} and α_{FA} , for which $\alpha_{T_2w} + \alpha_{FA} = \mathbf{1}$ at every voxel. The input velocity fields are weighted with the attention maps and combined to create a final velocity field v .

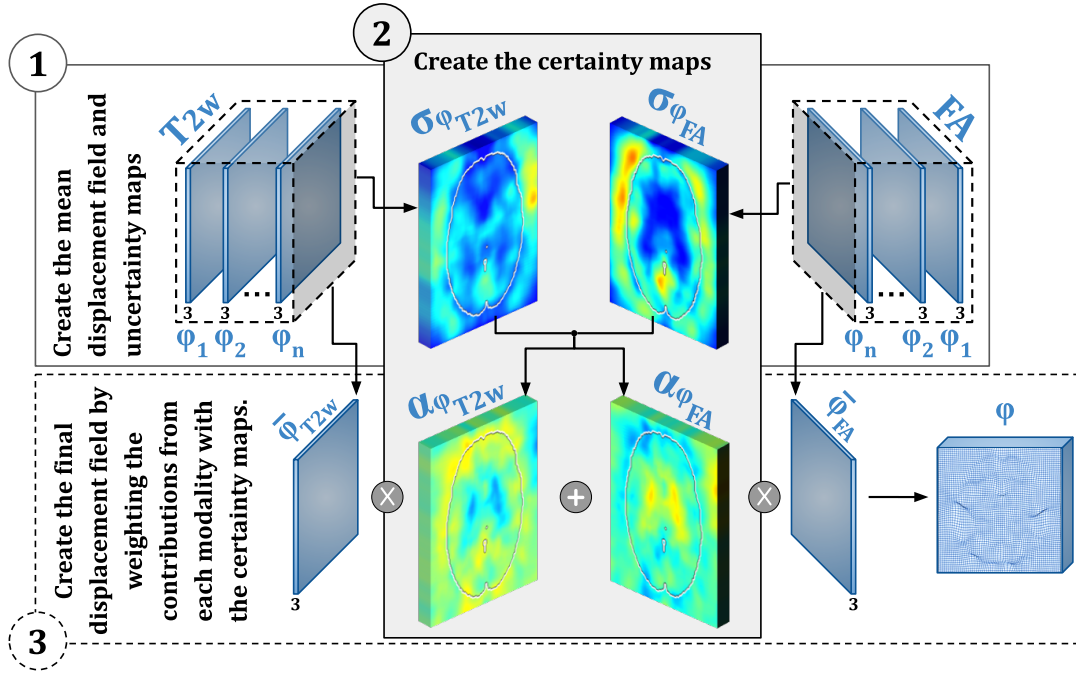


Figure 5.3: The construction of uncertainty-aware deformation field: ① Create modality-specific mean deformation fields $\bar{\varphi}_{T2w}$ and $\bar{\varphi}_{FA}$, and uncertainty maps $\sigma_{\varphi_{T2w}}$ and $\sigma_{\varphi_{FA}}$. ② Create modality specific certainty maps $\alpha_{\varphi_{T2w}}$ and $\alpha_{\varphi_{FA}}$ using equation 5.1. ③ Create the final deformation field by locally weighting the contributions from each modality with the certainty maps.

The architecture of our proposed *attention image registration network* is presented in Figure 5.4.

For each subject in our dataset, we employ the previously trained *registration-only networks* on either pairs of $T2w$ images (\mathbf{M}_{T2w} and \mathbf{F}_{T2w}) or FA maps (\mathbf{M}_{FA} and \mathbf{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 [377]. The activation maps of the final layer are concatenated with the subject’s moving images \mathbf{M}_{T2w} and \mathbf{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

$$v = v_{T2w} \odot \alpha_{T2w} + v_{FA} \odot \alpha_{FA} \quad (5.3)$$

where \odot represents the element-wise multiplication. Similar to the registration network, the velocity field v is put through an *exponentiation layer* to create the

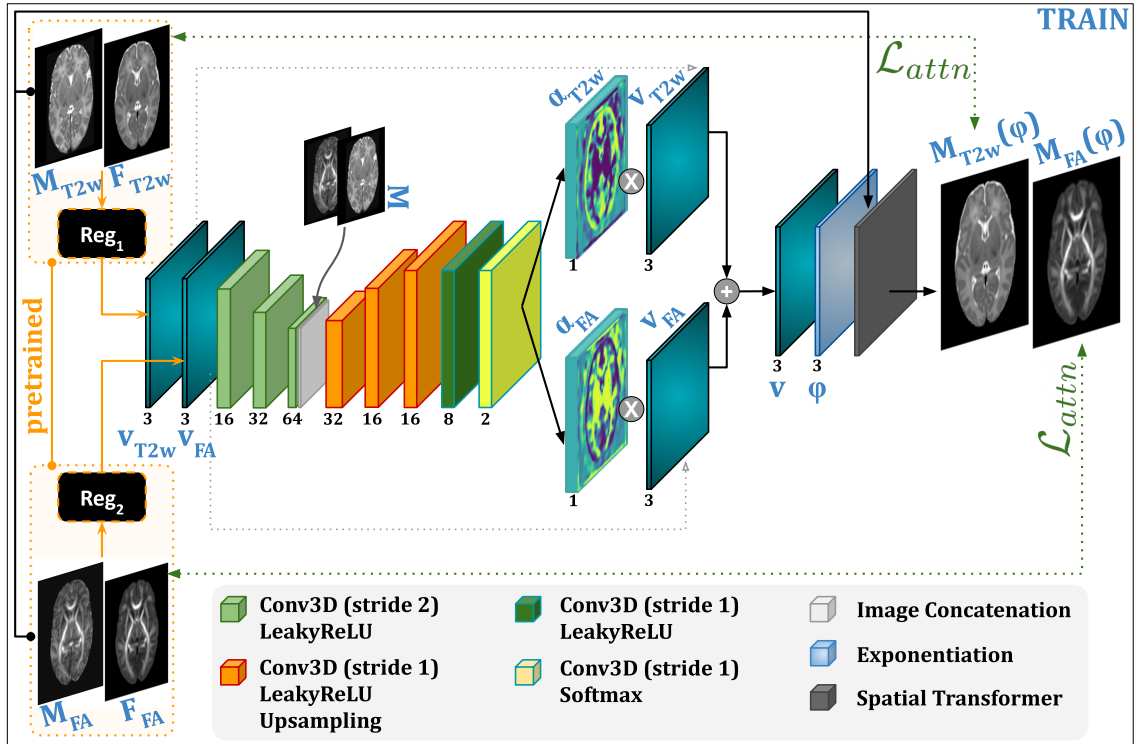


Figure 5.4: The proposed attention-based image registration network architecture, which uses as input subject- and modality-specific velocity fields (v_{T2w} and v_{FA}). The attention network outputs two 1-channel maps α_{T2w} and α_{FA} which are used to create a combined velocity field v . The velocity field v is transformed into a dense deformation field ϕ which warps the subject’s moving images (\mathbf{M}_{T2w} and \mathbf{M}_{FA}) into $\mathbf{M}_{T2w}(\phi)$ and $\mathbf{M}_{FA}(\phi)$. The network is trained to achieve good alignment between the warped images and the fixed atlas (\mathbf{F}_{T2w} and \mathbf{F}_{FA}). Note that the $T2w$ -only (Reg_1) and FA-only (Reg_2) networks are pre-trained.

combined deformation field ϕ , which is then used to warp the moving volumes \mathbf{M}_{T2w} and \mathbf{M}_{FA} .

Channel and spatial attention. 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 [309]), followed by a spatial attention module [314]. In total, we add 4 channel and spatial attention modules in the encoder part of the CVAE, and 5 modules in the decoder.

5.2.3 Training the networks

Training the *baseline* image registration networks. 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 deformation field φ which, together with the moving image \mathbf{M} , produces the warped image $\mathbf{M}(\varphi)$. During training, the optimizer aims to minimize both the KL divergence (equation 3.10), the BE regularisation penalty [68] (equation 2.13), and the reconstruction loss:

$$\begin{aligned} \mathcal{L}_{reg} = & \underbrace{KL[q_\omega(z|\mathbf{F}, \mathbf{M}) || p(z)]}_{\mathcal{D}_{KL}} + \lambda_{BE} \mathcal{L}_{BE}(\varphi) + \\ & \lambda \underbrace{\left(\lambda_{T2w} \mathcal{D}_{NCC}(\mathbf{F}^{T2w}(\varphi^{-\frac{1}{2}}), \mathbf{M}^{T2w}(\varphi^{\frac{1}{2}})) + \lambda_{FA} \mathcal{D}_{NCC}(\mathbf{F}^{FA}(\varphi^{-\frac{1}{2}}), \mathbf{M}^{FA}(\varphi^{\frac{1}{2}})) \right)}_{\mathcal{L}_{rec}} \end{aligned} \quad (5.4)$$

where λ , λ_{BE} , λ_{T2w} and λ_{FA} are hyperparameters. Throughout this work, $\lambda_{BE} = 0.01$ and $\lambda = 5000$, as proposed in [252]. \mathcal{L}_{BE} [68] was defined in equation 2.13, but we include it here for the sake of completion:

$$\begin{aligned} \mathcal{L}_{BE}(\varphi) = \sum_{\mathbf{x} \in \Omega} & \left[\left(\frac{\partial^2 \varphi(\mathbf{x})}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \varphi(\mathbf{x})}{\partial y^2} \right)^2 + \left(\frac{\partial^2 \varphi(\mathbf{x})}{\partial z^2} \right)^2 + \right. \\ & \left. 2 \left(\frac{\partial^2 \varphi(\mathbf{x})}{\partial xy} \right)^2 + 2 \left(\frac{\partial^2 \varphi(\mathbf{x})}{\partial xz} \right)^2 + 2 \left(\frac{\partial^2 \varphi(\mathbf{x})}{\partial yz} \right)^2 \right] \end{aligned} \quad (5.5)$$

Finally, \mathcal{D}_{NCC} is the symmetric NCC dissimilarity measure defined as:

$$\mathcal{D}_{NCC}(\mathbf{F}(\varphi^{-\frac{1}{2}}), \mathbf{M}(\varphi^{\frac{1}{2}})) = - \frac{\sum_{\mathbf{x} \in \Omega} (\mathbf{F}(\varphi^{-\frac{1}{2}}) - \bar{F}) \cdot (\mathbf{M}(\varphi^{\frac{1}{2}}) - \bar{M})}{\sqrt{\sum_{\mathbf{x} \in \Omega} (\mathbf{F}(\varphi^{-\frac{1}{2}}) - \bar{F})^2 \cdot \sum_{\mathbf{x} \in \Omega} (\mathbf{M}(\varphi^{\frac{1}{2}}) - \bar{M})^2}} \quad (5.6)$$

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

In equation 5.4, \mathcal{D}_{KL} aims to reduce the gap between the prior $p(z)$, defined as a multivariate unit Gaussian distribution $p(z) \sim \mathcal{N}(0, I)$, and the encoded distribution $q_\omega(z|\mathbf{F}, \mathbf{M})$. \mathcal{L}_{BE} regularizes the transformation φ by penalizing high bending energy, and \mathcal{L}_{rec} aims to reduce the reconstruction loss between the fixed image \mathbf{F} and warped image $\mathbf{M}(\varphi)$.

Using this setup, we train 2 single-modality baseline 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 multi-channel baseline models 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 8 baseline networks trained from scratch, until convergence (see the first row in Table 5.2).

Uncertainty-aware registration. The uncertainty-aware registration is achieved at inference-time, using the pre-trained baseline T_2w -only ($\lambda_{T2w} = 1.0$, $\lambda_{FA} = 0.0$) and FA-only ($\lambda_{T2w} = 0.0$, $\lambda_{FA} = 1.0$) networks. More specifically, equation 5.1

shows how the certainty maps $(\alpha_{\varphi_{T2w}}, \alpha_{\varphi_{FA}})$ are created from the modality-specific deformation fields. Note that in this case both channels have equal weights ($\lambda_{T2w} = \lambda_{FA} = 1.0$). To allow for different weightings of the $T2w$ and FA channels, we apply the following equations:

$$\alpha'_{\varphi_{T2w}} = \frac{\lambda_{T2w}\alpha_{\varphi_{T2w}}}{\lambda_{T2w}\alpha_{\varphi_{T2w}} + \lambda_{FA}\alpha_{\varphi_{FA}}} ; \alpha'_{\varphi_{FA}} = \frac{\lambda_{FA}\alpha_{\varphi_{FA}}}{\lambda_{T2w}\alpha_{\varphi_{T2w}} + \lambda_{FA}\alpha_{\varphi_{FA}}} \quad (5.7)$$

As $\alpha_{\varphi_{T2w}} + \alpha_{\varphi_{FA}} = \mathbf{1}$, when $\lambda_{T2w} = \lambda_{FA} = 1.0$, equation 5.7 reduces to: $\alpha'_{\varphi_{T2w}} = \alpha_{\varphi_{T2w}}$ and $\alpha'_{\varphi_{FA}} = \alpha_{\varphi_{FA}}$.

To summarize, for the uncertainty-aware multi-channel registration, we use the pre-trained $T2w$ -only and FA-only baseline networks to create modality-specific certainty maps, and we build the uncertainty-aware deformation fields by locally weighting the contributions from each modality with the certainty maps. We do this for the following sets of hyperparameters: $(\lambda_{T2w}, \lambda_{FA}) = \{(1.0, 0.1), (1.0, 0.175), (1.0, 0.25), (1.0, 0.5), (1.0, 0.75), (1.0, 1.0)\}$. This is summarized in the second row of Table 5.2.

Training the *channel + spatial attention registration networks.* The channel + spatial attention networks are trained similarly to the baseline networks, using the same loss function (equation 5.4), and the same hyperparameters: $\lambda_{BE} = 0.01$, $\lambda = 5000$, and $(\lambda_{T2w}, \lambda_{FA}) = \{(1.0, 0.1), (1.0, 0.175), (1.0, 0.25), (1.0, 0.5), (1.0, 0.75), (1.0, 1.0)\}$. This is summarized in the third row of Table 5.2.

Training the *attention-driven registration networks.* The attention-driven registration networks use as input the subject- and modality-specific velocity fields (v_{T2w} and v_{FA}) produced by the pre-trained baseline $T2w$ -only ($\lambda_{T2w} = 1.0$, $\lambda_{FA} = 0.0$) and FA-only ($\lambda_{T2w} = 0.0$, $\lambda_{FA} = 1.0$) networks (see Figure 5.4). During training, the optimizer aims to minimize the following loss function:

$$\mathcal{L}_{attn} = \lambda_{T2w} \mathcal{D}_{\text{NCC}}(\mathbf{F}^{T2w}(\varphi^{-\frac{1}{2}}), \mathbf{M}^{T2w}(\varphi^{\frac{1}{2}})) + \lambda_{FA} \mathcal{D}_{\text{NCC}}(\mathbf{F}^{FA}(\varphi^{-\frac{1}{2}}), \mathbf{M}^{FA}(\varphi^{\frac{1}{2}})) \quad (5.8)$$

where \mathcal{D}_{NCC} is the symmetric NCC dissimilarity measure, and φ is the obtained through *scaling and squaring* layers applied to the velocity field obtained through equation 5.3. Similarly to before, we train the attention-driven registration networks with the following set of hyperparameters: $(\lambda_{T2w}, \lambda_{FA}) = \{(1.0, 0.1), (1.0, 0.175), (1.0, 0.25), (1.0, 0.5), (1.0, 0.75), (1.0, 1.0)\}$. This is summarized in the final row of Table 5.2.

To summarize, we train 8 *baseline* image registration networks (2 single-channel and 6 multi-channel), 6 *channel+spatial attention* networks, and 6 proposed *attention-based* networks. The 6 *uncertainty-aware* experiments are obtained from the single-channel *baseline* networks at inference time. We train all of the 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 [378] with a base learning rate of 10^{-6} and a maximum learning rate of 10^{-3} , and an L_2 weight decay (L_2 penalty) factor of 10^{-5} . All networks were implemented in

Model	T2w-only			T2w+FA				FA-only	
	λ_{FA}	0	.1	.175	.25	.5	.75	1	1
	λ_{T2w}	1	1	1	1	1	1	1	0
<i>baseline</i>		✓	✓	✓	✓	✓	✓	✓	✓
<i>uncertainty</i>			✓	✓	✓	✓	✓	✓	
<i>ch+sp</i>			✓	✓	✓	✓	✓	✓	
<i>attention</i>			✓	✓	✓	✓	✓	✓	

Table 5.2: Single- and multi-channel experiment setups used in this study, for different values of hyperparameters λ_{FA} and λ_{T2w} .

PyTorch (v1.10.2), with TorchIO (v0.18.73) [206] for data preprocessing (intensity normalisation) and loading, and training was performed on a 12 GB Titan XP. Average inference times were: 0.16s/sample for the *baseline* networks, 0.31s/sample for the proposed *attention-based* networks, and 0.63s/sample for the *channel+spatial attention* networks.

5.3 Results

5.3.1 Quantitative evaluation

To validate which of the 26 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 [170]: cGM, WM, ventricles, hippocampi and amygdala. Additionally, a WM structure called the internal capsule (IC) was manually segmented on the 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 [364]) were calculated between the warped labels and the atlas labels.

Cortical gray matter vs. internal capsule. First, we looked at how the models performed based on two tissue types (the cGM and the IC). We chose these two 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. Dice scores are summarised in Figure 5.5, while average surface distances are summarised in Figure 5.6. In both figures, the first column (shaded in light pink) shows the values for the initial affine alignment, while the second and last columns show the T_2w -only and the FA-only baseline 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. For visualisation purposes, each figure consists of 4 plots: the first three compare the baseline networks with the: a) *channel+spatial attention*, b)

uncertainty-aware, and c) our *proposed attention*, respectively. The last plot shows all models together (baseline vs. all).

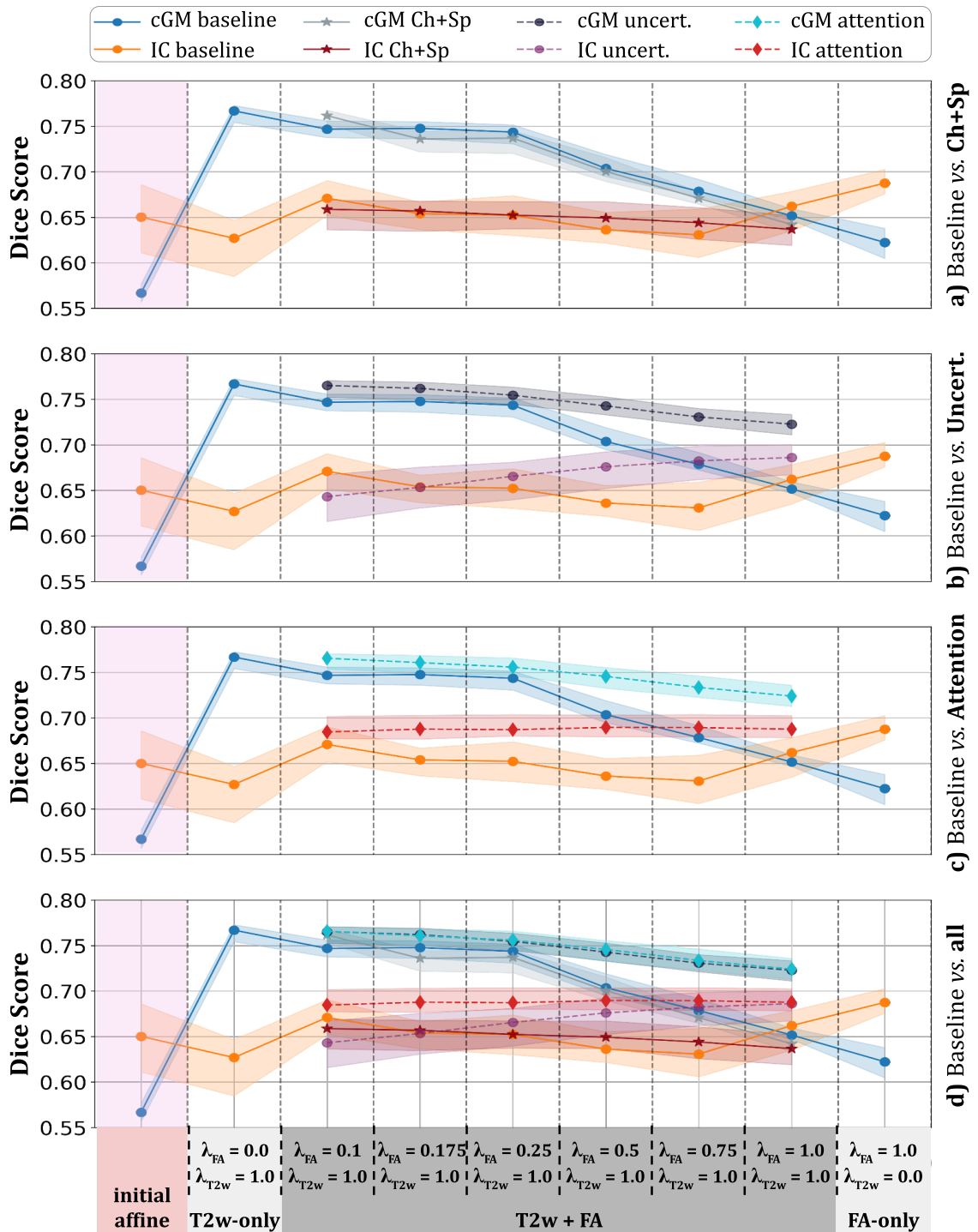


Figure 5.5: Line plots showing median Dice scores for cGM and IC structures, with the first column showing their initial affine alignment. The first three plots compare the baseline networks with the: a) *channel+spatial* attention, b) *uncertainty-aware*, and c) our *proposed attention*, respectively. The last plot (d) aggregates all of these results into one figure.

The best overall performance in terms of Dice scores and average surface dis-

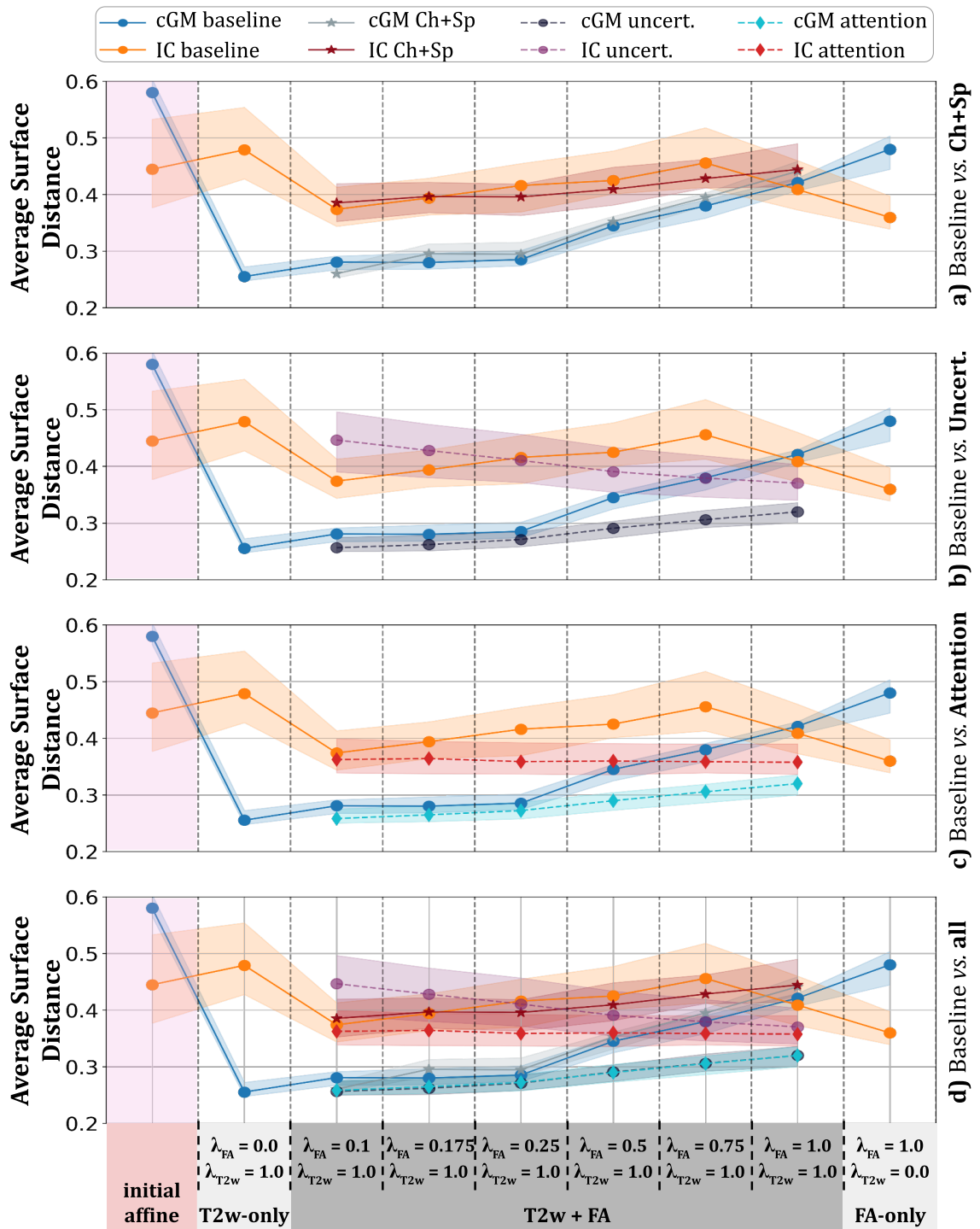


Figure 5.6: Line plots showing median average surface distances for cGM and IC structures, with the first column showing their initial affine alignment. The first three plots compare the baseline networks with the: a) *channel+spatial* attention, b) *uncertainty-aware*, and c) our *proposed attention*, respectively. The last plot (d) aggregates all of these results into one figure.

tances is obtained by our proposed *attention* model for $\lambda_{T2w} = 1.0$ and $\lambda_{FA} = 0.1$ (third column, Figures 5.5 and 5.6), 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 ($\lambda_{T_2w} = 1.0$ 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*. Similarly, the *uncertainty-aware* model yields good results when compared to our proposed *attention* for the cGM structure, but cannot align the IC well, obtaining results significantly lower than all of the other models.

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 *baseline* 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 outperforms the multi-channel *baseline* registration networks, and this improvement is statistically significant for all values of λ_{FA} .

Multiple structural labels. Table 5.3 shows the results of 7 of our models for all available tissue types (cGM, WM, ventricles, hippocampi and amygdala, and IC), with Dice scores showing in the top half of the table (marked by the **DS** label on the last column), and average surface distances in the second half of the table (marked by the **ASD** label on the last column). A two-sample, two-sided paired t-test with a significance level of 5% was used to compare pairs of the trained models. Statistically significant differences ($p\text{-value} < 0.05$) are reported in Table 5.3 in terms of best overall score (highlighted in bold), and best/worst amongst the multi-channel models (highlighted in green/red, respectively). The initial affine alignment is shown in the first row of each type of score (DS or ASD). The table also contains the results for four baseline networks: T_2w -only, FA-only, T_2w +FA when they are both weighed the same: $\lambda_{T_2w} = \lambda_{FA} = 1.0$, and T_2w +FA when $\lambda_{T_2w} = 1.0$ and $\lambda_{FA} = 0.1$. The other multi-channel models, *channel+spatial attention*, *uncertainty-aware*, and our proposed *attention*, are shown in the last three rows of the DS and ASD scores, respectively. All models marked with the T_2w +wFA label on the first column of the table are trained with the lowest weight on the FA maps ($\lambda_{T_2w} = 1.0$ and $\lambda_{FA} = 0.1$). The $\lambda_{FA} = 0.1$ was chosen here because it showcases the best overall results in the previous section.

Our proposed *attention* 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 the 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.

Model		cGM	WM	Ventricles	Amygdala	IC	
initial affine		.567±.02	.700±.03	.631±.05	.746±.05	.642±.07	
baseline	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	
	T2w+FA	.653±.01	.766±.01	.742±.03	.782±.02	.655±.03	
T2w+wFA	baseline	.747±.01	.826±.02	.775±.02	.808±.02	.669±.03	ASD
	ch+sp	.761±.01	.841±.01	.791±.01	.814±.02	.656±.03	
	uncert	.763±.01	.842±.01	.792±.01	.809±.02	.638±.04	
	attention	.763±.01	.842±.01	.793±.02	.816±.02	.683±.03	
initial affine		.582±.04	.409±.04	.508±.1	.310±.08	.479±.1	
baseline	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	
	T2w+FA	.419±.02	.317±.02	.324±.06	.266±.04	.417±.06	
T2w+wFA	baseline	.279±.01	.218±.02	.264±.04	.223±.04	.383±.05	ASD
	ch+sp	.262±.01	.198±.01	.248±.04	.209±.03	.390±.05	
	uncert	.261±.02	.197±.01	.246±.05	.224±.04	.456±.08	
	attention	.260±.02	.197±.01	.248±.04	.212±.03	.370±.05	

Table 5.3: Mean (\pm standard deviation) Dice scores (DS) in the first half of the table and average surface distances (ASD) in the second half of the table. All results are on the test set, with the initial affine alignment shown on the first rows of each score. First, the single channel baseline networks (T_2w -only and FA-only) are shown, followed by the multi-channel $\lambda_{T_2w} = \lambda_{FA} = 1.0$ baseline network. The following 4 rows showcase the multi-channel, $\lambda_{T_2w} = 1.0$ and $\lambda_{FA} = 0.1$, for the baseline, the *channel+spatial* attention, the *uncertainty-aware* attention and our *proposed attention* model, respectively. Overall best scores are highlighted in bold (p -value < 0.05). The green shading highlights the model which performed best amongst the multi-channel models (p -value < 0.05), while the red shading points to the multi-channel models which performed worst.

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 *baseline* $T_2w+\mathbf{w}FA$ model). Similarly, the *uncertainty-aware* models performed well for the structural labels (cGM, WM and ventricles), but had the poorest scores for the IC (lowest amongst all the multi-channel models, and compared to the initial affine alignment).

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 baseline models always performed worse than the *attention-driven* models. In fact, the T_2w+FA network, where $\lambda_{T_2w} = \lambda_{FA} = 1.0$, 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.

5.3.2 Qualitative results

Visualisation of attention maps. Figure 5.7 shows average attention maps from 10 neonatal subjects scanned around 40 weeks PMA for two of our *attention-driven* models (when $\lambda_{FA} = \lambda_{T_2w} = 1.0$ and when $\lambda_{T_2w} = 1.0, \lambda_{FA} = 0.1$). The first two columns of Figure 5.7 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_{T_2w} attention maps (in atlas space) for the 2 models. Specifically, for each subject $j \in [1, 10]$ and model $m \in \{\textit{attention with } \lambda_{FA} = \lambda_{T_2w} = 1.0, \textit{attention with } \lambda_{FA} = 0.1, \lambda_{T_2w} = 1.0\}$, we obtained attention maps $\alpha_{T_2w}^{jm}$ and α_{FA}^{jm} , and averaged them across the subjects: $\bar{\alpha}_{T_2w}^m = \frac{1}{10} \sum_{j=1}^{10} \alpha_{T_2w}^{mj}$ and $\bar{\alpha}_{FA}^m = \frac{1}{10} \sum_{j=1}^{10} \alpha_{FA}^{mj}$. Figure 5.7 shows the $\bar{\alpha}_{T_2w}^m$ maps only, as their FA counterparts are $\bar{\alpha}_{FA}^m = 1 - \bar{\alpha}_{T_2w}^m$. We can observe that the $\bar{\alpha}_{T_2w}$ 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_{T_2w} is close to zero in the area of the main white matter tracts in both cases, suggesting that the FA channel is used in this area for the registration task. Moreover, this qualitative finding explains the overall steadiness of the IC results across different λ_{FA} values shown in Figures 5.5 and 5.6.

In fact, this is quite evident in Figure 5.8, where both $\bar{\alpha}_{T_2w}$ and $\bar{\alpha}_{FA}$ maps are shown for all pairs of hyperparameters: $(\lambda_{T_2w}, \lambda_{FA}) = \{(1.0, 0.1), (1.0, 0.175), (1.0, 0.25), (1.0, 0.5), (1.0, 0.75), (1.0, 1.0)\}$, in both mid-brain axial and coronal slices. As λ_{FA} increases from 0.1 to 1.0, the $\bar{\alpha}_{T_2w}$ maps become less pronounced in the cGM regions. This is shown in Figure 5.8 through the green arrow which points at a cGM region in both axial and coronal slices. When $\lambda_{FA} = 0.1$, these regions are close to 1, but they become less reliant on the T_2w channel with increasing the FA global weight. On the other hand, the FA channel remains stable for the IC structure, and this is shown on the α_{FA} maps in Figure 5.8 with the cyan arrows and ovals. The

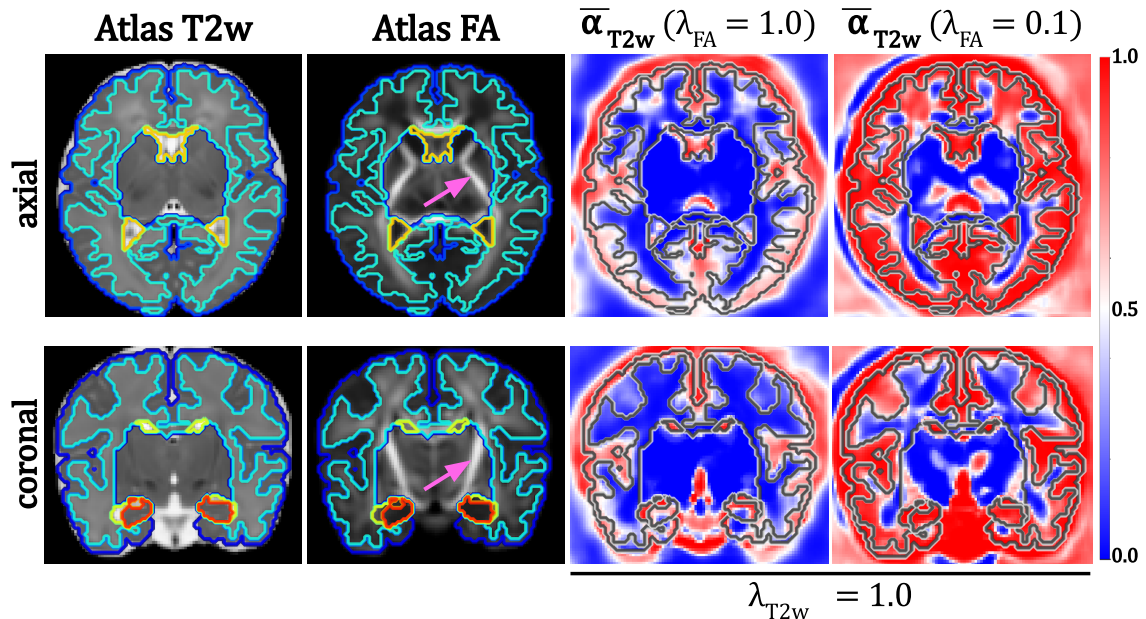


Figure 5.7: Mid-brain axial and coronal slices of both T_2w and FA fixed images (first two columns), together with average α_{T_2w} attention maps for the proposed *attention* multi-channel registration network with $\lambda_{FA} = \lambda_{T_2w} = 1.0$, and with $\lambda_{T_2w} = 1.0$, $\lambda_{FA} = 0.1$, respectively. 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.

α_{FA} map values remain close to 1 inside the IC for all λ_{FA} hyperparameters.

Comparison with uncertainty-aware maps. Figure 5.9 compares the average attention maps produced by our proposed *attention* model with the *uncertainty-aware* model, for $\lambda_{FA} = \lambda_{T_2w} = 1.0$. The *uncertainty-aware* maps are much noisier than our proposed model, and this is especially pronounced in the IC region (see cyan arrows in the α_{FA} maps of Figure 5.9). We hypothesize that this is the reason why the IC is poorly aligned for the *uncertainty-aware* model. For the sake of completion, we add in the first column the attention maps produced by networks trained with higher Gaussian smoothing (with $\sigma = 5$ mm instead of $\sigma = 1$ mm). Indeed, the maps are smoother, but at this resolution the cGM is poorly aligned (with Dice scores below 0.65 and average surface distances above 0.45). Similarly, the IC is not as well aligned as when $\sigma = 1$ mm, with values significantly poorer than our proposed *attention* network (Dice scores below 0.7 and average surface distances above 0.4).

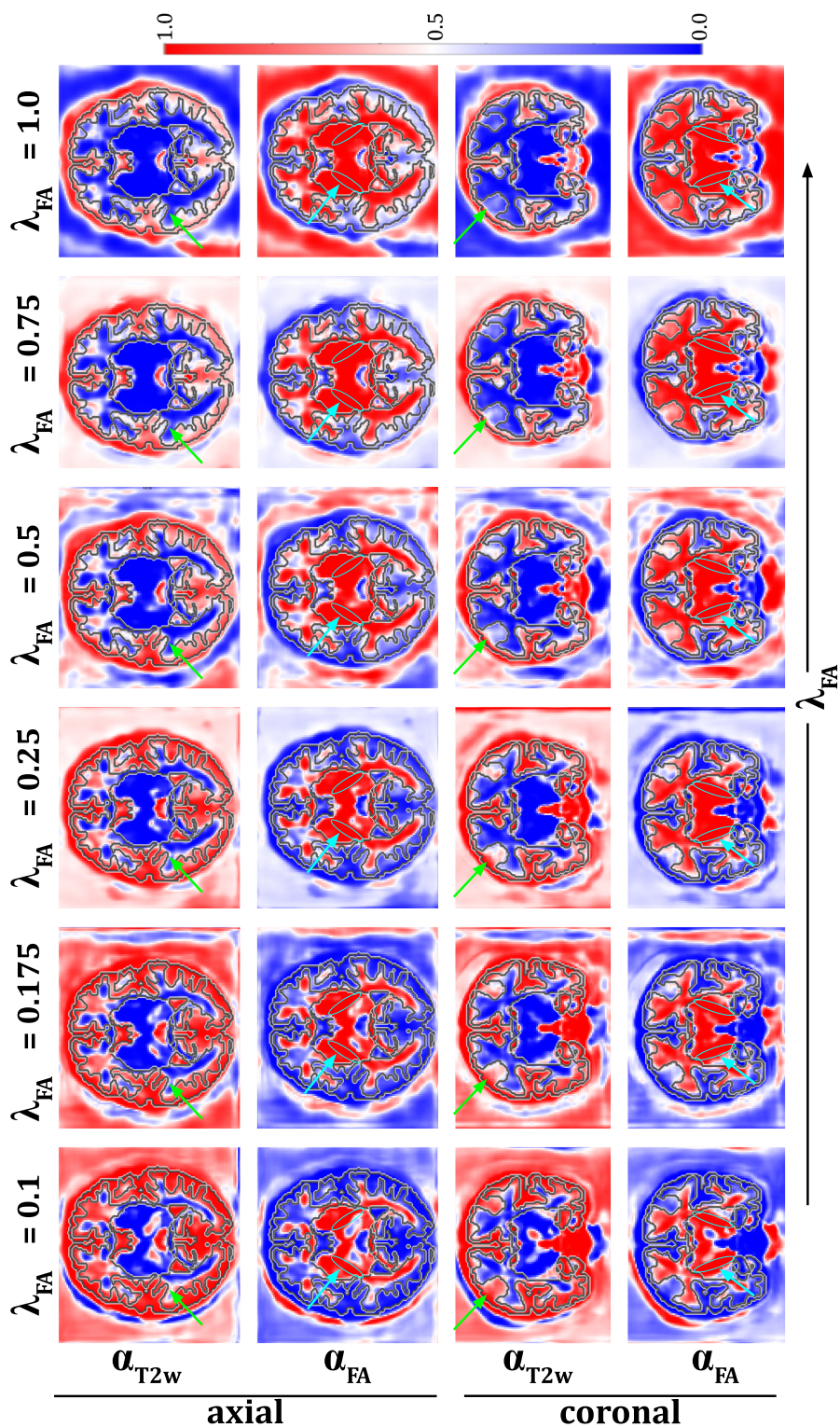


Figure 5.8: Mid-brain axial and coronal slices of average α_{T2w} and α_{FA} attention maps for $\lambda_{FA} = \{0.1, 0.175, 0.25, 0.5, 0.75, 1.0\}$. Contour lines of the boundaries between cGM, WM, ventricles and hippocampi and amygdala are overlaid on top. The green arrows point to regions of cGM (in both axial and coronal slices) of the α_{T2w} maps which becomes less dependent on the T_{2w} channel as λ_{FA} increases. The cyan arrows and ovals point to regions of the IC (in both axial and coronal slices) of the α_{FA} maps which remain dependent on the FA channel as λ_{FA} changes value.

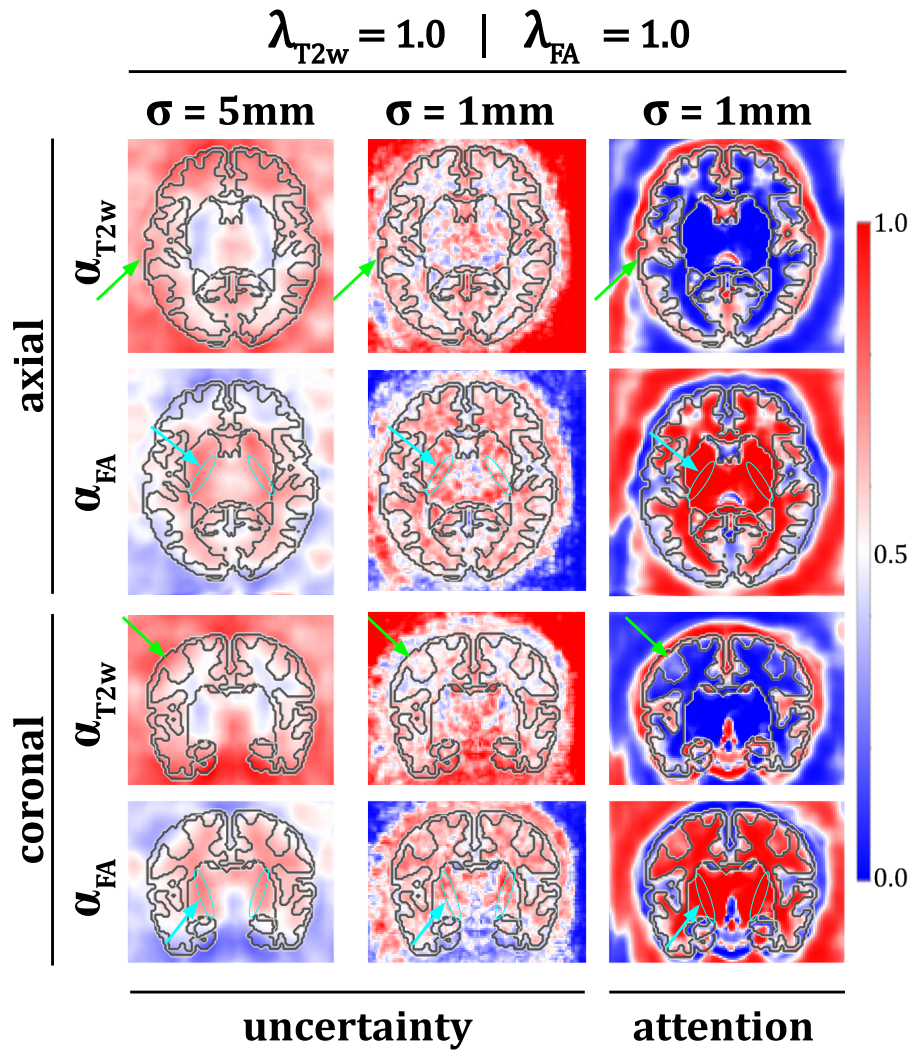


Figure 5.9: Mid-brain axial and coronal slices of average α_{T2w} and α_{FA} attention maps for $\lambda_{FA} = \lambda_{T2w} = 1.0$. The first two columns show the attention maps derived with the *uncertainty-aware* model, while the last column shows the proposed *attention* maps. The first column shows the attention maps when using an increased kernel size for the Gaussian smoothing layer ($\sigma = 5$ mm), while the last two columns show the default parameter chosen for this study ($\sigma = 1$ mm). Contour lines of the boundaries between cGM, WM, ventricles and hippocampi and amygdala are overlaid on top. The green arrows point to regions of cGM (in both axial and coronal slices) of the α_{T2w} maps, while the cyan arrows and ovals point to regions of the IC (in both axial and coronal slices) of the α_{FA} maps.

5.4 Discussion and future work

In this work we presented 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 development of reliable methods for neonatal brain MR image registration holds significant clinical relevance. For example, accurate image registration is an important step in building age-dependent templates and spatiotemporal atlases [379, 373] which are used as reference for identifying normal and abnormal brain structures, or establishing normative developmental ranges. Moreover, studies involving brain morphometry, which have shown promising results in infants at risk of developing cognitive or sensorimotor impairments [380, 379], require accurate registration and spatial normalisation of the subjects in a common space. For instance, our proposed attention model has the potential to become a key component in studying neurodevelopmental outcomes using tensor-based morphometry [381], as this requires accurate alignment of each subject to a common template space.

Deploying such MR image registration models in a clinical setting is not without challenges, and there are several potential gaps that need to be addressed. In particular, to develop accurate image registration models it is important to have access to diverse datasets of neonatal brain MR images. However, there is often a scarcity of annotated neonatal imaging data, particularly with reliable registration information as the ground truth. At the same time, the current implementation relies on a well-curated dataset in which the training data has undergone specific preprocessing steps, such as resizing, adjusting spatial orientation, and applying affine registration to a common template space. Our work can potentially serve as a step within a more complex pipeline where all these prerequisites must be fulfilled beforehand.

Our study has several limitations. First, our experiments lack a comparison with more classical image registration methods, such as ANTs [60] or NiftyReg [113], where the fusion of different channels could be performed either with or without certainty maps calculated from normalized gradients correlated to structural content [51]. Second, the available segmentation maps used to quantify the accuracy of the label propagation through the predicted deformation fields were obtained through an automated process (i.e., Draw-EM). More specifically, our quantitative evaluation relies on the assumption that the available tissue maps are correct, without being able to compare against inter-observer measures performed by medical raters. At the same time, this study was focused on combining information from T_2w neonatal scans with DWI-derived FA maps, restricting its applicability to higher-order data,

such as DTI. Moreover, the CVAE employed in this study was trained using a latent space size of 32, based on the original paper by Krebs *et al.* [253], without analysing how different values (higher or lower) might affect the results. At the same time, we kept the same latent space size for both single- and multi-channel data, which might have an impact on the capacity of the networks. In terms of our proposed attention network, in this study we only analysed the scenario in which the attention maps are single-channel, *i.e.*, they yield a scalar value per-voxel, thus weighting each spatial dimension of the velocity field equally. In future work, we aim to investigate if a 3-channel attention map (1 channel for each dimension of the velocity field) further improves our results. Finally, we used only one white matter label, namely the IC, for showcasing the importance of using microstructural data. In the future, we aim to explore multiple white matter structures, such as the external capsule and the corpus callosum.

Diffusion tensor driven deep learning image registration

Motivation

Registration of DT images has the potential to better align WM structures than using structural MRI only. Moreover, unlike the scalar-valued FA maps, DTI enables the inclusion of fiber orientation at each voxel.

Contribution

An extension of the previously proposed attention-based multi-channel deep learning image registration framework to deal with higher-order DT image data, by accounting for the change in orientation of diffusion tensors induced by the predicted deformation field.

Publications

- Grigorescu, I. et al. (2020). *Diffusion Tensor Driven Image Registration: A Deep Learning Approach*. WBIR 2020. LNCS (Springer)
🔗 https://doi.org/10.1007/978-3-030-50120-4_13

Code available at:

🔗 github.com/irinagrigorescu/attentionneonatalmriregistration

6.1 Introduction

Medical image registration is a vital component of a large number of clinical applications. For example, image registration is used to track longitudinal changes occurring in the brain. However, most applications in this field rely on a single modality, without taking into account the rich information provided by other modalities. Although T_2 w MRI scans provide good contrast between different brain tissues, they do not have knowledge of the extent or location of white matter tracts. Moreover, during early life, the brain undergoes dramatic changes, such as cortical folding and myelination, processes which affect not only the brain’s shape, but also the MRI tissue contrast.

In order to establish correspondences between images acquired during the neonatal period, we propose a deep learning image registration framework which combines T_2 w and DW MRI scans. More specifically, this study extends the previously proposed multi-channel image registration network (see Section 5) with layers capable of dealing with higher-order DT data. As the fixed images, we use the same 36 weeks old neonatal structural (T_2 w) atlas as described in Section 5, and instead of FA maps, we use the corresponding microstructural DTI atlas [51] of the same age. Throughout this work we use MRI brain scans acquired as part of the dHCP project [11] for the moving images.

6.2 Methods

6.2.1 Data acquisition and preprocessing

The MRI data used in this study was collected as part of the dHCP project [11] and details about the data acquisition can be found in Section 1.2.3. More specifically, we used 414 3D T_2 w and DTI volumes of neonates born between 23 – 42 weeks GA and scanned at term-equivalent age (37 – 45 weeks PMA). As preprocessing steps, we first affinely pre-registered the data to a common 36 weeks gestational age atlas space [51] using the MIRTk software toolbox [68], and then we resampled both structural and microstructural volumes to be 1 mm isotropic resolution. To obtain the DT maps, we used the `DWI2TENSOR` [382] command available in the MRtrix3¹ toolbox [375], and we performed skull-stripping using the available dHCP brain masks [148]. Finally, we cropped the resulting images to a $128 \times 128 \times 128$ size.

The dataset partition for training, validating and testing our networks is described in Table 6.1, and was kept similar to the study shown in Chapter 5. We

¹<https://mrtrix.readthedocs.io/>

used the validation set to inform us about our models’ performance during training, and we report all of our results on the test set.

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)

Table 6.1: 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).

6.2.2 Network architectures

Baseline image registration network. Let F , M represent the fixed (target) and the moving (source) MR volumes, respectively, and let φ be the deformation field. In this work, the focus is on T_2w MRI volumes (F^{T2w} and M^{T2w}) which are single channel data, and DTI volumes (F^{DTI} and M^{DTI}) which are 6 channel data. The moving images M^{T2w} and M^{DTI} are acquired from the same subjects, while the fixed images F^{T2w} and F^{DTI} are the 36 weeks old neonatal atlas [51].

The overall architecture of the proposed network, which is similar to the baseline network described in Section 5 (see Figure 5.1), is shown in Figure 6.1. During training, our model uses pairs of T_2w and/or DT images to learn a velocity field v , which is transformed into a topology-preserving deformation field φ through *scaling and squaring* layers. The *Spatial Transformer* layer [181] receives as input the predicted field φ and the moving scalar-valued T_2w image, and outputs the warped and resampled image. A similar process is necessary to warp the moving DT image, with a few extra steps which are explained in the next subsection.

Attention-based image registration network. For our attention-based multi-channel image registration network, we employ the previously proposed CNN (see Chapter 5, Figure 5.4) which uses pairs of modality-specific velocity fields as input, and outputs a combined velocity field which aims to align both structural (T_2w) and microstructural (DTI) data simultaneously. The network learns the 1-channel attention maps α_{T2w} and α_{DTI} , for which $\alpha_{T2w} + \alpha_{DTI} = \mathbf{1}$ at every voxel. The input velocity fields (v_{T2w} and v_{DTI}) are weighted with the attention maps and combined to create a final velocity field:

$$v = v_{T2w} \odot \alpha_{T2w} + v_{DTI} \odot \alpha_{DTI} \quad (6.1)$$

The architecture of our proposed *attention image registration network* is presented in Figure 6.2.

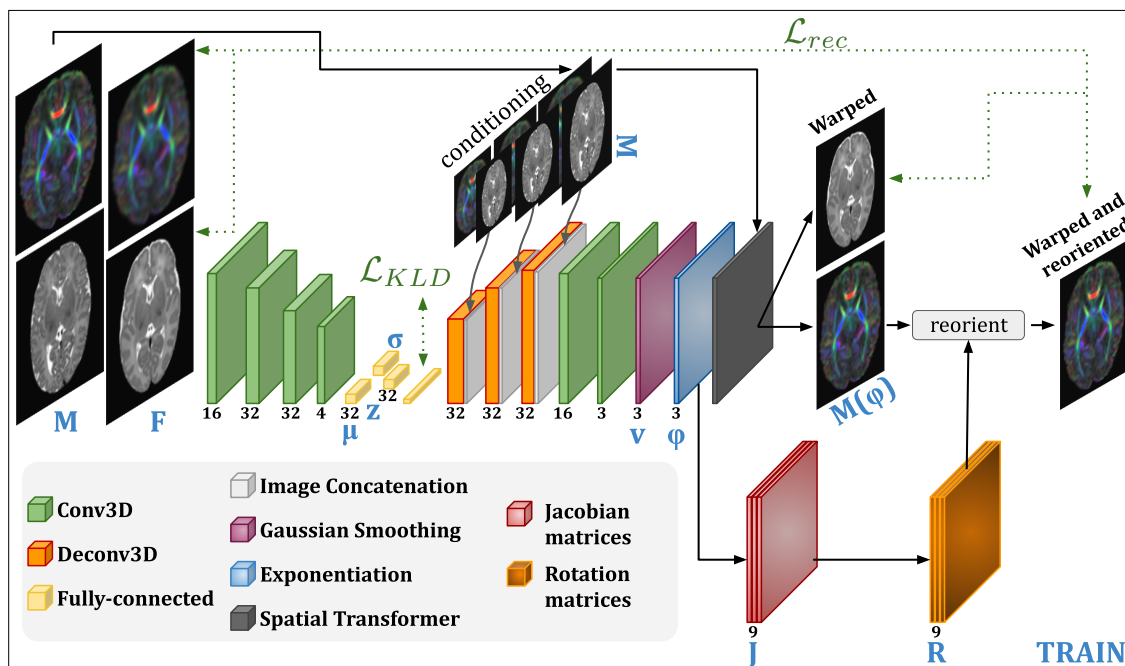


Figure 6.1: The diffusion tensor driven multi-channel image registration network at training time. Jacobian matrices J are calculated at each spatial location for the currently predicted deformation field φ and then used to calculate the rotation matrices R through polar decomposition. The final DT moved image is obtained after tensor reorientation. The reconstruction loss (\mathcal{L}_{rec}) is computed between the fixed atlases F^{T2w} and F^{DTI} and the warped structural $M^{T2w}(\varphi)$ and the warped microstructural image with reoriented tensors $M^{DTI}(\varphi)$. Note that the $\mathbf{9}$ value shown above the J and R matrices represents the 9-channel data, as both the Jacobian matrices J and the rotation matrices R are calculated for each voxel. Thus, for a deformation field φ of shape $N_B \times 3 \times 128 \times 128 \times 128$, the J and R matrices will be $N_B \times 9 \times 128 \times 128 \times 128$ (N_B is the number of batches).

6.2.3 Tensor reorientation

Registration of DT images is not as straightforward to perform as scalar-valued data. When transforming the latter, the intensities in the moving image are interpolated at the new locations determined by the deformation field φ and copied to the corresponding location in the target image space. However, after interpolating DT images, the diffusion tensors need to be reoriented to remain anatomically correct [105]. This is done by applying a rotation matrix R to each resampled diffusion tensor \mathbf{D} , such that: $\mathbf{D}' = R\mathbf{D}R^T$.

When the transformation is non-linear, such as in our case, the reorientation matrix can be computed at each point in the deformation field φ through a polar decomposition of the local Jacobian matrix. This factorisation transforms the non-singular matrix J into a unitary matrix R (the pure rotation) and a positive-semidefinite Hermitian matrix P , such that $J = RP$ [106]. The rotation matrices R are then used to reorient the tensors without changing the local microstructure.

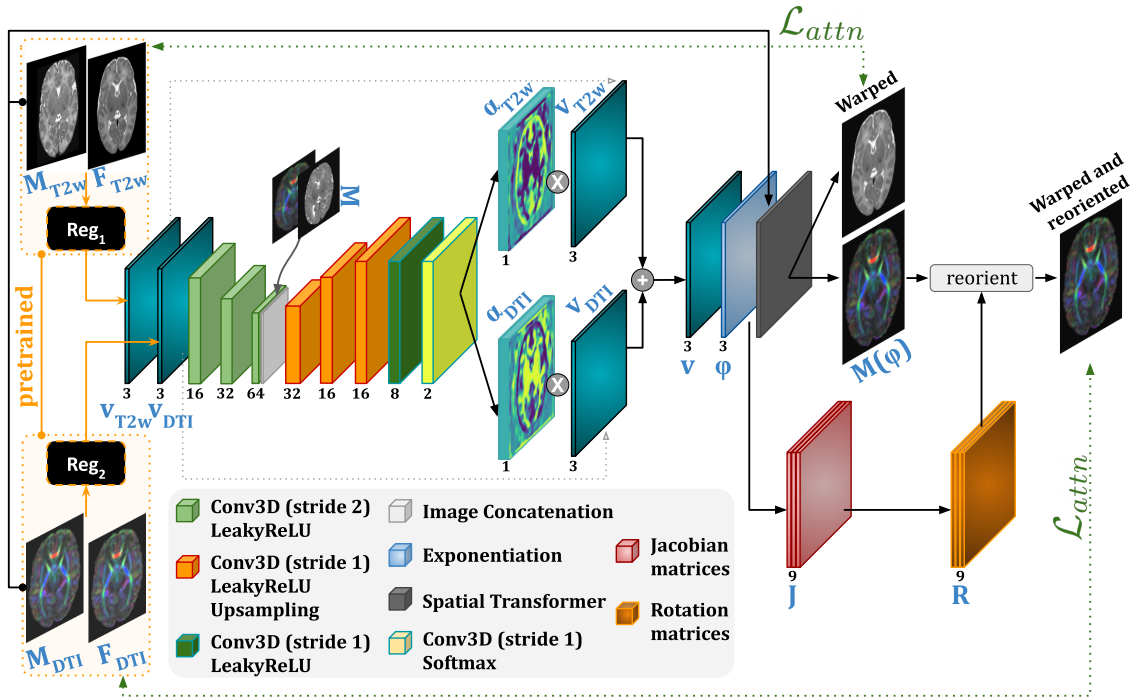


Figure 6.2: The proposed diffusion tensor driven multi-channel attention image registration network at training time. The overall architecture is kept similar to the study presented in Chapter 5 (Figure 5.4), with the added components necessary for DT reorientation. The loss (\mathcal{L}_{attn}) is computed between the fixed atlases F^{T2w} and F^{DTI} and the warped structural $M^{T2w}(\varphi)$ and the warped microstructural image with reoriented tensors $M^{DTI}(\varphi)$.

This is known as the *finite strain* strategy [105].

In our proposed framework, a Jacobian matrix J is calculated from the predicted deformation field φ at each spatial location \mathbf{x} (see Figures 6.1 and 6.2), using equation 2.14. For the sake of completion, we include the equation here:

$$\mathbf{JAC}(\varphi(\mathbf{x})) = \begin{bmatrix} \frac{\partial \varphi_x(\mathbf{x})}{\partial x} & \frac{\partial \varphi_x(\mathbf{x})}{\partial y} & \frac{\partial \varphi_x(\mathbf{x})}{\partial z} \\ \frac{\partial \varphi_y(\mathbf{x})}{\partial x} & \frac{\partial \varphi_y(\mathbf{x})}{\partial y} & \frac{\partial \varphi_y(\mathbf{x})}{\partial z} \\ \frac{\partial \varphi_z(\mathbf{x})}{\partial x} & \frac{\partial \varphi_z(\mathbf{x})}{\partial y} & \frac{\partial \varphi_z(\mathbf{x})}{\partial z} \end{bmatrix} \quad (6.2)$$

6.2.4 Training the networks

Training the *baseline* image registration network. We train our **baseline** networks using a combination of the KL divergence (equation 3.10), the BE regu-

larisation penalty [68] (equation 2.13), and the reconstruction loss:

$$\mathcal{L}_{reg} = \mathcal{D}_{\text{KL}} + \lambda_{BE} \mathcal{L}_{BE}(\varphi) + \underbrace{\lambda \left(\lambda_{T2w} \mathcal{D}_{\text{NCC}}(\mathbf{F}^{T2w}(\varphi^{-\frac{1}{2}}), \mathbf{M}^{T2w}(\varphi^{\frac{1}{2}})) + \lambda_{DTI} \mathcal{D}_{\text{EDS}}(\mathbf{F}^{DTI}(\varphi^{-\frac{1}{2}}), \mathbf{M}^{DTI}(\varphi^{\frac{1}{2}})) \right)}_{\mathcal{L}_{rec}} \quad (6.3)$$

In the equation 6.3 shown above, \mathcal{D}_{KL} aims to reduce the gap between the prior $p(z)$, defined as a multivariate unit Gaussian distribution $p(z) \sim \mathcal{N}(0, I)$, and the encoded distribution $q_{\omega}(z | \mathbf{F}, \mathbf{M})$. \mathcal{L}_{BE} regularizes the transformation φ by penalizing high bending energy, and \mathcal{L}_{rec} aims to reduce the reconstruction loss between the fixed image \mathbf{F} and warped image $\mathbf{M}(\varphi)$.

In short, we use the same reconstruction loss for the structural data as we did in the study presented in Chapter 5 (equation 5.6). For the microstructural data, in order to encourage a good alignment between the DT images, we employ one of the most commonly used diffusion tensor similarity measures, known as the Euclidean distance squared [61] which was previously presented in equation 2.26. For the sake of completion, we include it here:

$$\mathcal{D}_{\text{EDS}}(\mathbf{F}(\varphi^{-\frac{1}{2}}), \mathbf{M}(\varphi^{\frac{1}{2}})) = \sum_{\mathbf{x} \in \Omega} \|\mathbf{F}(\varphi^{-\frac{1}{2}}) - \mathbf{M}(\varphi^{\frac{1}{2}})\|_C^2 \quad (6.4)$$

where the **euclidean distance** between **two tensors** \mathbf{D}_1 and \mathbf{D}_2 is:

$$\|\mathbf{D}_1 - \mathbf{D}_2\|_C = \sqrt{\text{Tr}((\mathbf{D}_1 - \mathbf{D}_2)^2)} \quad (6.5)$$

Training the attention-driven registration network. Similar to the FA study presented in Chapter 5, the attention-driven registration networks use as input the subject- and modality-specific velocity fields (v_{T2w} and v_{DTI}) produced by the pre-trained **baseline** T_2w -only and DTI-only networks. During training, the optimizer aims to minimize the following loss function:

$$\mathcal{L}_{attn} = \lambda_{T2w} \mathcal{D}_{\text{NCC}}(\mathbf{F}^{T2w}(\varphi^{-\frac{1}{2}}), \mathbf{M}^{T2w}(\varphi^{\frac{1}{2}})) + \lambda_{DTI} \mathcal{D}_{\text{EDS}}(\mathbf{F}^{DTI}(\varphi^{-\frac{1}{2}}), \mathbf{M}^{DTI}(\varphi^{\frac{1}{2}})) \quad (6.6)$$

where φ is the field obtained through *scaling and squaring* layers applied to the velocity field defined in equation 6.1.

Particularities of training with DTI data. One major drawback of training with DTI data is the use of a voxel-wise singular value decomposition (SVD) for calculating the reorientation matrices. This has proven to be time consuming, with a computation time of ~ 3 s/sample. Moreover, unlike the previous T_2w +FA study where we were able to use the global symmetric NCC dissimilarity measure for both structural and microstructural data, the Euclidean distance squared [61] is one order of magnitude smaller than the values obtained with NCC. We empirically found

that multiplying the \mathcal{D}_{EDS} by 10 brings the values closer in range. At the same time, to better understand the right balance between the two losses without spending the time consuming task of training multiple 3D models, we have performed hyperparameter tuning on 2D mid-brain axial slices of the dataset. These experiments are summarised in Table 6.2 for both the *baseline* and the proposed *attention* models. More specifically, we trained the 2D networks for a range of hyperparameter values: $(\lambda_{T2w}, \lambda_{DTI}) = \{(1.0, 0.1), (1.0, 0.5), (1.0, 1.0), (1.0, 1.5), (1.0, 2.0), (1.0, 3.0)\}$, as well as DTI-only ($\lambda_{T2w} = 0.0, \lambda_{DTI} = 1.0$).

Model		DTI-only		T2w+DTI				
	λ_{DTI}	1	.1	.5	1	1.5	2	3
	λ_{T2w}	0	1	1	1	1	1	1
2D baseline		✓	✓	✓	✓	✓	✓	✓
2D attention			✓	✓	✓	✓	✓	✓
3D baseline		✓			✓	✓	✓	
3D attention					✓	✓	✓	

Table 6.2: Single- and multi-channel 2D and 3D experiment setups used in this study, for different values of hyperparameters λ_{DTI} and λ_{T2w} .

To summarize, we train 7 *baseline* 2D image registration networks (1 model for DTI-only and 6 multi-channel models), and 6 *attention-based* multi-channel networks (see the **2D baseline** and **2D attention** rows in Table 6.2). For the 3D networks, we train 4 *baseline* 3D image registration networks (1 model for DTI-only and 3 multi-channel), and 3 proposed *attention-based* network with the hyperparameters that we found to perform well in the 2D experiments (see the **3D baseline** and **3D attention** rows in Table 6.2).

We train all of the models until convergence, with a maximum of 150 epochs for 2D training and a maximum of 300 epochs for 3D training. We also employ the Adam optimizer with its default parameters ($\beta_1=.9$ and $\beta_2=.999$), and use the models which performed best on the validation set. Similar to Chapter 5, we employ a decaying cyclical learning rate scheduler [378], but with a smaller base learning rate of $5 \cdot 10^{-8}$ and a maximum learning rate of $5 \cdot 10^{-4}$, and a larger L_2 weight decay (L_2 penalty) factor of 10^{-4} . This is because we observed that training with the original parameters caused an accumulation of large derivatives which resulted in the model being unstable and incapable of learning or producing viable predictions. All networks were implemented in PyTorch (v1.10.2), with TorchIO (v0.18.73) [206] for data preprocessing (intensity normalisation) and loading, and training was performed on a 12 GB Titan XP.

6.3 Results

6.3.1 Quantitative evaluation

The quantitative evaluation is carried out 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 [170]: cGM, WM, ventricles, and hippocampi and amygdala. Additionally, a WM structure called the IC was manually segmented on the 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 [364]) were calculated between the warped labels and the atlas labels.

Hyperparameter tuning on 2D mid-brain axial slices. For the 2D experiments described in Table 6.2, we calculated Dice scores and average surface distances of 4 out of the 5 available labels (we excluded hippocampi and amygdala as it was not present in the mid-brain axial slices used). A two-sample, two-sided paired t-test with a significance level of 5% was used to compare the models, and statistically significant differences ($p\text{-value} < 0.05$) are summarised in Table 6.3. Similar to our previous 3D experiments from Chapter 5, the T_2w -only model performs best on cGM and WM structures, with the microstructural data (DTI-only in this case) performing best for the IC. This is visible in Table 6.3 for both Dice scores and average surface distances (the best overall values are in bold).

For the multi-channel models, as λ_{DTI} increases from 0.1 to 3.0, the overall tendency is a decrease in performance for the cGM or WM tissue types, with the opposite happening for the IC and ventricles (better scores for higher λ_{DTI}). In fact, the ventricles are aligned best with either the T_2w+DTI or the proposed attention-based T_2w+DTI models, outperforming the T_2w -only network. We also find that the multi-channel **attention** models always outperform the multi-channel **baseline** models, or perform similarly well for the ventricles, for the same λ_{DTI} .

As these networks are trained with 2D mid-brain axial slices, some brain features may not be present to guide the alignment. However, the overall trend shows that higher values of λ_{DTI} will cause a decrease in the alignment of cGM and WM, while increasing the alignment of ventricles and IC. For this reason, we choose to train the 3D models with $\lambda_{DTI} \in [1.0, 1.5, 2.0]$, and the next section will focus on our 3D experiments.

Model (λ_{DTI})		cGM	WM	Ventricles	IC	
affine		.611±.02	.678±.03	.697±.07	.578±.08	
T2w		.786±.01	.821±.01	.816±.03	.577±.07	
DTI		.715±.01	.762±.02	.822±.03	.675±.03	
T2w + DTI	0.1	.779±.01	.811±.01	.815±.03	.600±.05	DS
	0.5	.780±.01	.806±.01	.825±.03	.616±.05	
	1.0	.782±.01	.812±.01	.824±.03	.630±.05	
	1.5	.778±.01	.803±.02	.828±.03	.643±.04	
	2.0	.774±.01	.805±.02	.828±.03	.651±.04	
	3.0	.774±.01	.801±.02	.834±.03	.658±.04	
attention	0.1	.786±.01	.820±.01	.817±.03	.600±.06	ASD
	0.5	.785±.01	.819±.01	.826±.03	.650±.04	
	1.0	.785±.01	.818±.01	.829±.02	.660±.03	
	1.5	.785±.01	.817±.01	.829±.03	.664±.03	
	2.0	.784±.01	.814±.01	.830±.03	.667±.03	
	3.0	.779±.01	.811±.02	.831±.03	.670±.04	
affine		.579±.06	.537±.06	.471±.16	.549±.18	
T2w		.258±.02	.281±.03	.208±.04	.542±.17	
DTI		.365±.03	.377±.05	.234±.07	.374±.06	
T2w + DTI	0.1	.268±.02	.295±.03	.208±.05	.492±.12	ASD
	0.5	.266±.02	.302±.04	.201±.04	.469±.11	
	1.0	.262±.02	.295±.03	.201±.04	.452±.11	
	1.5	.268±.02	.309±.04	.191±.04	.425±.09	
	2.0	.273±.02	.304±.04	.188±.04	.417±.08	
	3.0	.274±.02	.313±.04	.187±.04	.393±.08	
attention	0.1	.258±.02	.281±.03	.201±.04	.500±.15	ASD
	0.5	.258±.02	.284±.03	.201±.04	.415±.08	
	1.0	.259±.02	.284±.03	.192±.04	.398±.08	
	1.5	.259±.02	.287±.03	.19±.05	.390±.07	
	2.0	.259±.02	.291±.03	.189±.04	.386±.07	
	3.0	.260±.02	.294±.03	.189±.04	.380±.06	

Table 6.3: Mean (\pm standard deviation) Dice scores (DS) and average surface distances (ASD) of the 2D T_2w and DTI experiments. The initial affine alignment is shown first, followed by the single channel **baseline** networks (T_2w -only and DTI-only), the multi-channel T_2w +DTI **baseline** networks, and ending with the proposed multi-channel **attention** models. Overall best scores are highlighted in bold (p -value < 0.05). The green shading highlights the model which performed best amongst the multi-channel models (p -value < 0.05), while the red shading points to the multi-channel models which performed worst.

Quantitative evaluation of our 3D models. For our 3D experiments, we looked at the performance of 7 DTI-based registration models: 4 *baseline* models (1 DTI-only and 3 T_2w +DTI models with different λ_{DTI} values), and 3 proposed T_2w +DTI *attention-based* networks (see Table 6.2 for reference). We compared the performance of these models to the T_2w -only, FA-only and attention-based T_2w +FA model scores obtained in Chapter 5. Both multi-channel experiments, *i.e.*, **baseline** T_2w +DTI and **attention-based** T_2w +DTI, were trained with $\lambda_{DTI} \in [1.0, 1.5, 2.0]$. Similar to the 2D experiments, a two-sample, two-sided paired t-test with a significance level of 5% was used to compare pairs of the trained 3D models. Table 6.4 summarizes these results, where the T_2w -only, FA-only and attention-based T_2w +FA (with $\lambda_{FA} = 0.1$) models are highlighted in gray to show that the values are obtained from the previously trained networks.

The proposed **attention-based** network (see **A** T_2w +DTI rows in Table 6.4) outperforms the **baseline** T_2w +DTI models for all $\lambda_{DTI} \in [1.0, 1.5, 2.0]$. In fact, for $\lambda_{DTI} = 1.0$, the latter obtains both Dice scores and average surface distances significantly worse ($p\text{-value} < 0.05$) than all of the other multi-channel models, as highlighted by the red shading. In terms of the IC, the proposed attention-based T_2w +DTI models, as well as the DTI-only model, perform similarly well ($p\text{-value} > 0.05$) to the FA-only and the previously proposed attention-based T_2w +FA (with $\lambda_{FA} = 0.1$) network, in terms of both Dice scores and average surface distances, and regardless of the λ_{DTI} used.

Our proposed attention T_2w +DTI model experienced a drop in performance in terms of aligning the cGM and the WM structures. More specifically, for $\lambda_{DTI} = 2.0$, the model obtained a decrease in Dice scores of 0.014 for cGM and 0.019 for WM, respectively, as well as an increase in average surface distances of 0.019 for cGM and 0.027 for WM, respectively. Lowering the λ_{DTI} from 2.0 to 1.0 increased the performance of the attention models in terms of aligning cGM and WM structures, while the scores obtained for the IC decreased, but not significantly. We hypothesize that, for the T_2w +FA model, the attention maps will prioritize using the T_2w channel in the cGM ribbon as the FA maps have little information in this area. On the other hand, DT images are rich in information across the entire brain, and in this case the T_2w +DTI model is more likely to choose from both channels when aligning this structure (see also Figure 6.5).

Model	(λ_{DTI})	cGM	WM	Ventricles	Amygdala	IC	
affine		.567±.02	.700±.03	.631±.05	.746±.05	.642±.07	
T2w-only		.763±.01	.844±.02	.797±.02	.803±.02	.614±.04	
FA-only		.621±.02	.756±.02	.676±.04	.769±.03	.686±.03	
DTI-only		.679±.02	.767±.02	.757±.03	.786±.02	.682±.03	
T2w+DTI	1.0	.706±.01	.746±.03	.755±.03	.739±.02	.648±.03	DS
T2w+DTI	1.5	.711±.01	.776±.02	.773±.03	.746±.02	.650±.03	
T2w+DTI	2.0	.713±.01	.766±.03	.766±.03	.752±.02	.660±.04	
A T2w+FA		.763±.01	.842±.01	.793±.02	.816±.02	.683±.03	
A T2w+DTI	1.0	.758±.01	.834±.02	.799±.02	.809±.02	.677±.03	
A T2w+DTI	1.5	.752±.01	.829±.02	.798±.02	.809±.02	.681±.03	
A T2w+DTI	2.0	.749±.01	.825±.02	.796±.02	.810±.02	.681±.03	
affine		.582±.04	.409±.04	.508±.1	.310±.08	.479±.1	
T2w-only		.259±.02	.193±.02	.242±.05	.233±.04	.498±.09	
FA-only		.477±.04	.319±.02	.433±.09	.276±.05	.374±.05	
DTI-only		.375±.02	.307±.03	.296±.05	.250±.04	.376±.05	
T2w+DTI	1.0	.340±.02	.355±.04	.290±.05	.371±.04	.443±.07	ASD
T2w+DTI	1.5	.334±.02	.316±.03	.276±.05	.359±.04	.431±.06	
T2w+DTI	2.0	.330±.02	.331±.04	.280±.05	.341±.04	.424±.07	
A T2w+FA		.260±.02	.197±.01	.248±.04	.212±.03	.37±.05	
A T2w+DTI	1.0	.266±.02	.206±.02	.238±.04	.221±.04	.388±.06	
A T2w+DTI	1.5	.274±.02	.215±.02	.241±.04	.221±.04	.382±.05	
A T2w+DTI	2.0	.278±.02	.220±.02	.242±.04	.219±.04	.382±.05	

Table 6.4: Mean (\pm standard deviation) Dice scores (DS) and average surface distances (ASD) of the 3D T_2w and DTI experiments. The initial affine alignment is shown first, followed by the single channel **baseline** networks (T_2w -only, FA-only and DTI-only), the multi-channel T_2w +DTI **baseline** networks (with $\lambda_{DTI} \in [1.0, 1.5, 2.0]$), and ending with the proposed multi-channel **attention** models (T_2w +FA with $\lambda_{FA} = 0.1$, and T_2w +DTI with $\lambda_{DTI} \in [1.0, 1.5, 2.0]$). Overall best scores are highlighted in bold (p -value < 0.05). The green shading highlights the model which performed best amongst the multi-channel models (p -value < 0.05), while the red shading points to the multi-channel models which performed worst. **A** was used as shorthand notation for “attention”.

Evaluation of white matter alignment. To better understand how our proposed **attention-based** T_2w+DTI networks perform compared to the other models, we carried out an evaluation of the accuracy of WM alignment, using two metrics proposed in Adluru *et al.* [383]. First, the cross correlation (CC) between two scalar-valued maps is defined as [383]:

$$\text{CC} = \frac{\sum_v \mathbf{F}_{trace}^{DTI}(v) \mathbf{W}_{trace}^{DTI}(v)}{\sqrt{\sum_v \mathbf{F}_{trace}^{DTI}(v) \mathbf{F}_{trace}^{DTI}(v) \sum_v \mathbf{W}_{trace}^{DTI}(v) \mathbf{W}_{trace}^{DTI}(v)}} \quad (6.7)$$

where v is the index over all voxels, \mathbf{F}_{trace}^{DTI} and \mathbf{W}_{trace}^{DTI} are the maps corresponding to the fixed DTI atlas and the warped DTI test subject, with each spatial location containing the trace of the tensors. The higher the values obtained, the higher the similarity between the two maps in terms of the overall diffusivity in the tissue.

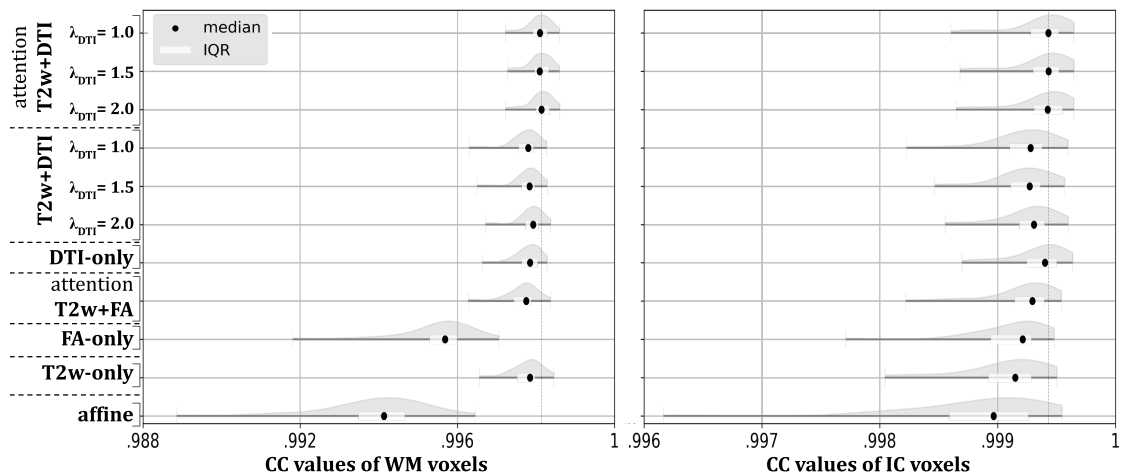
The second metric, known as the overlap of eigenvalue-eigenvector pairs (OVL), is able to investigate the alignment of diffusion tensors, with higher values representing better alignment. The OVL between two tensors is defined as [383, 384]:

$$\text{OVL} = \frac{\sum_{i=1}^3 \lambda_i^{\mathbf{F}} \lambda_i^{\mathbf{W}} (\varepsilon_i^{\mathbf{F}} \cdot \varepsilon_i^{\mathbf{W}})^2}{\sum_{i=1}^3 \lambda_i^{\mathbf{F}} \lambda_i^{\mathbf{W}}} \quad (6.8)$$

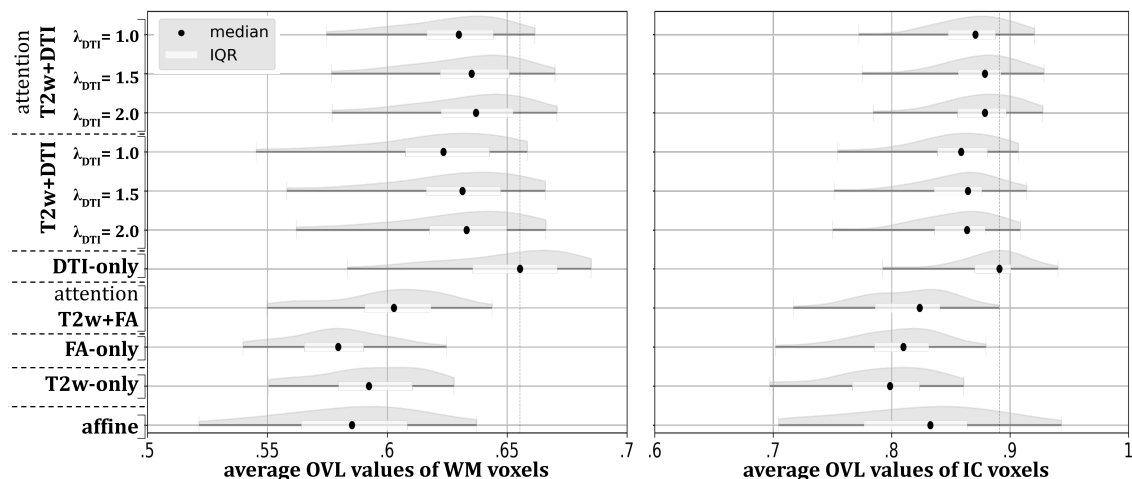
where $\lambda^{\mathbf{F}}$, $\varepsilon^{\mathbf{F}}$ are the eigenvalues and eigenvectors of the fixed DTI atlas volume (\mathbf{F}^{DTI}), while the $\lambda^{\mathbf{W}}$, $\varepsilon^{\mathbf{W}}$ are the eigenvalues and eigenvectors of the warped microstructural image with reoriented tensors $\mathbf{M}^{DTI}(\varphi)$. In equation 6.8, ‘ \cdot ’ is the dot product, and as the eigenvectors have unit norm, $\varepsilon_i^{\mathbf{F}} \cdot \varepsilon_i^{\mathbf{W}}$ represents a measure of the angle between the two corresponding vectors.

This evaluation is carried out on the same test dataset of 30 subjects, for 10 of our models: 7 multi-channel models (the **baseline** and the **attention** T_2w+DTI with $\lambda_{DTI} \in [1.0, 1.5, 2.0]$, and the **attention** T_2w+FA with $\lambda_{FA} = 0.1$ models), and 3 single-channel models (the **baseline** T_2w -only, FA-only and DTI-only models). For each model and each subject in our test dataset, we compute both the CC and the average OVL scores between the warped DT image with reoriented tensors and the fixed DTI atlas. More specifically, the predicted deformation fields for each type of model are used to warp and reorient the subject-specific DT images. The results are summarised in Figure 6.3, where on the left we show values computed across WM voxels (using the atlas WM label map), while on the right we show values calculated across voxels within the IC (using the atlas IC label map).

The results are plotted as violin plots for each of the 10 models, as well as the initial affine alignment. A two-sample, Wilcoxon signed rank test with 5% significance level was performed to test the null hypothesis of same distribution for different pairs of the trained models-derived CC and OVL scores. In terms of the CC scores, shown in Figure 6.3 a, the proposed **attention** T_2w+DTI model with $\lambda_{DTI} = 2.0$ performs best, obtaining significantly better ($p\text{-value} < 0.05$) values than all the other models, in terms of WM voxels. For the IC voxels, all **attention** T_2w+DTI models performed similarly well, obtaining significantly better results than all the other models. This suggests that using the T_2w and DTI modalities together helps with aligning the overall diffusivity in the tissue.



(a) Cross correlation scores of the WM voxels (left) and the IC voxels (right) between the warped 30 test subjects and the fixed atlas using the trace of the tensors.



(b) Average overlap of eigenvalue-eigenvector scores of the WM voxels (left) and the IC voxels (right) between the warped 30 test subjects and the fixed atlas.

Figure 6.3: CC scores (a) and average OVL values (b) of WM and IC voxels shown as violin plots for the initial affine alignment and 10 of our models: 3 single-channel models (T_2w -only, FA-only and DTI-only) and 7 multi-channel models (T_2w +DTI with $\lambda_{DTI} \in [1.0, 1.5, 2.0]$ with and without **attention**, as well as the previously proposed **attention** T_2w +FA with $\lambda_{FA} = 0.1$ model).

The average OVL values, which look at the directional components of the diffusion tensor, are shown in Figure 6.3 b. For both the WM and IC voxels, the best overall scores are obtained by the DTI-only model (see Figure 6.3 b). Interestingly, the FA-only model obtains lower OVL scores than the initial affine alignment for both WM and IC structures. Similarly, but only for the voxels within the IC, the T_2w -only and the **attention** T_2w +FA models obtain lower OVL scores than the initial affine alignment.

Using DTI helps with aligning the underlying WM structures and this is backed by the results shown in Figure 6.3 b, where all of the models which use DTI data have significantly higher OVL values (p -value < 0.05) when compared with the other models or the initial affine alignment. Moreover, this experiment shows that using **attention** when combining the T_2w and DTI data is better than not using attention, as the average OVL values in both the WM and IC voxels are significantly higher (p -value < 0.05) when compared to the **baseline** T_2w +DTI for the same λ_{DTI} . Finally, the **attention** T_2w +DTI model with $\lambda_{DTI} = 2.0$ obtains the closest average OVL values to the DTI-only model.

6.3.2 Qualitative results

Visualisation of 2D attention maps. Figure 6.4 shows mid-brain axial average attention maps from 10 neonatal subjects scanned around 40 weeks PMA for the *attention*-driven model trained with increasing values of λ_{DTI} , ranging from 0.1 to 3.0. We can observe that for the lowest $\lambda_{DTI} = 0.1$, the α_{T_2w} is covering the brain almost entirely, with the DTI map having little to no effect in training. As λ_{DTI} increases, the DTI modality has increasing importance (with values above 0.5 in the α_{DTI} maps) in WM regions. Moreover, this qualitative finding explains the increasing Dice scores and decreasing average surface distances for the IC structure as λ_{DTI} increases in value (see Table 6.3).

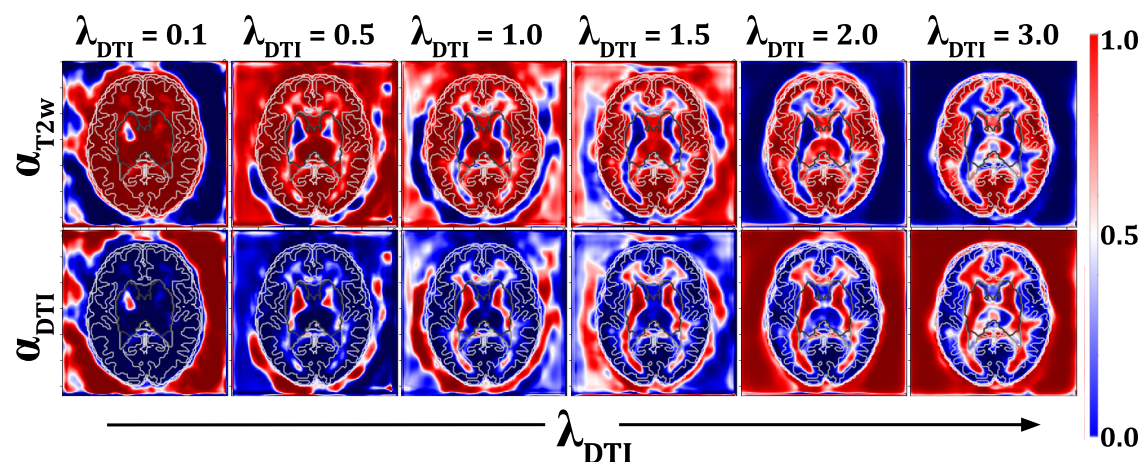


Figure 6.4: Average α_{T_2w} and α_{DTI} attention maps for the 2D *attention* multi-channel registration network for increasing values of λ_{DTI} , ranging from 0.1 to 3.0.

Visualisation of 3D attention maps. Using the same 10 neonatal subjects scanned around 40 weeks PMA, we computed average attention maps for the proposed 3D *attention*-driven model trained with $\lambda_{DTI} \in [1.0, 1.5, 2.0]$. These are shown in Figure 6.5, together with the maps produced by the previously proposed **attention** T_2w+FA with $\lambda_{FA} = 0.1$. Rows 1 and 3 show the structural α_{T2w} mid-brain axial and coronal slices, while rows 2 and 4 show the microstructural α_{DTI} and α_{FA} axial and coronal maps, respectively.

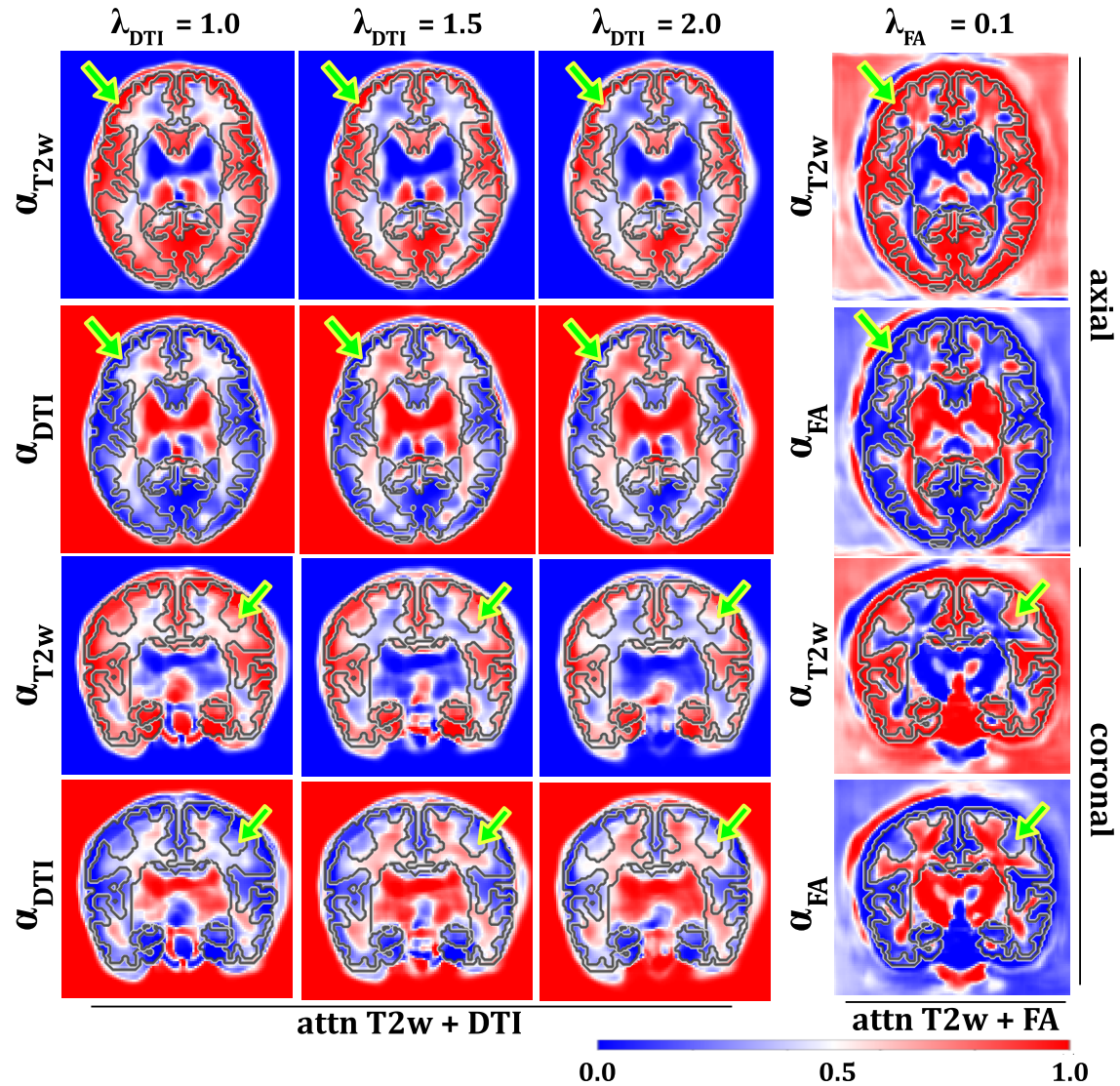


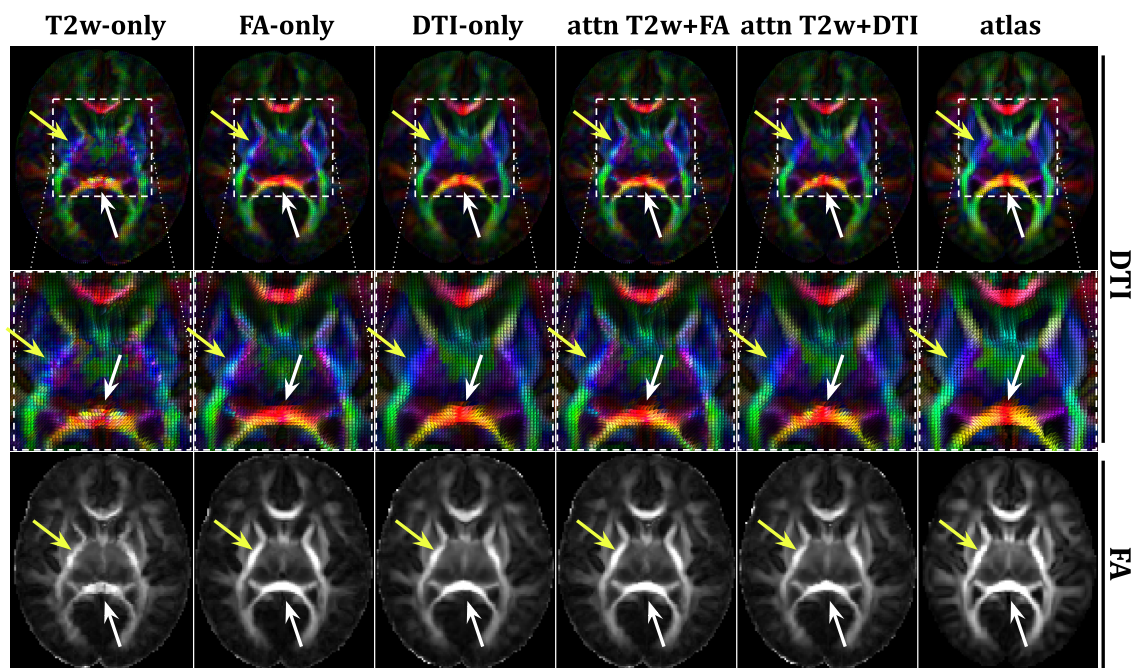
Figure 6.5: Average α_{T2w} and α_{DTI} attention maps for the 3D *attention* multi-channel registration network for $\lambda_{DTI} \in [1.0, 1.5, 2.0]$, as well as average α_{T2w} and α_{FA} attention maps for the 3D *attention* multi-channel registration network for $\lambda_{FA} = 0.1$. The green arrows point to regions in the cGM.

Unlike the 2D experiments, the α_{T2w} maps are less pronounced in the cGM regions (see Figure 6.4 *vs.* Figure 6.5). In fact, a similar conclusion can be drawn when comparing the attention T_2w+DTI maps with the attention T_2w+FA maps in Figure 6.5. More specifically, the α_{T2w} attention maps for the proposed attention T_2w+DTI model show that the cGM region is not as well delineated as in the

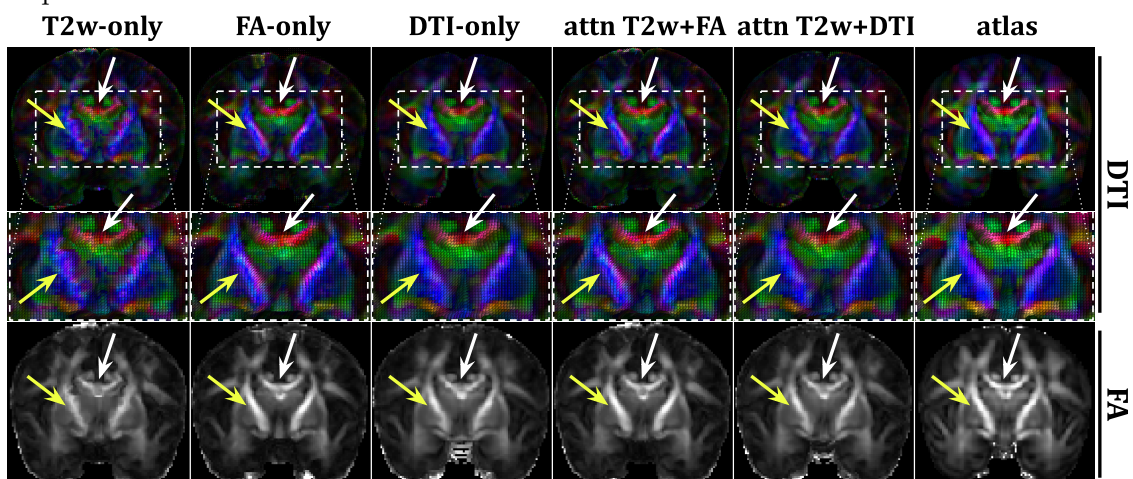
T_2w +FA case (*i.e.*, showcasing high values for the T_2w channel and low for the FA maps in the cGM areas), with the maps (in both axial and coronal views) *spilling over* onto the cortical ribbon. We hypothesize that this is due to FA images having very little contrast in the cGM regions, whereas DTI contains information across the entire brain.

Visualising average DTI maps. Finally, using the same 10 neonates, we looked at average DT maps. Figure 6.6 shows both mid-brain axial and coronal slices of the average DTI maps, as well as their corresponding FA maps (generated with the `TENSOR2METRIC MRtrix3` [375] command). The last column shows the DTI atlas and its respective FA map. In both the axial and coronal views, the white arrows point to the corpus callosum, an area which is strikingly different in DT orientations in the T_2w -only models when compared to the other models. Similarly, the yellow arrows point to regions of the IC, which, again, are more disorderly on the T_2w -only model. This is also highlighted in the generated FA maps, where the highlighted regions are evidently less coherent for the T_2w -only model.

In general, all the models using DTI as a single or an extra channel show an overall more consistent DT orientations in the average maps. The most striking visual difference between the average maps which were produced with the help of DT images and the ones which used FA (or only T_2w) is in the IC, as pointed out by the yellow arrows in the axial view (Figure 6.6 a). This is backed by our results from Figure 6.3 b, where average OVL values of IC voxels for models which use FA are significantly lower than any of the models which employ DTI, and even significantly lower than the initial affine alignment.



(a) Mid-brain axial slices of average DTI maps for 5 of our trained models (we used $\lambda_{DTI} = 2.0$, and $\lambda_{FA} = 0.1$ for the **attention** models), together with the fixed DTI atlas and FA map. Yellow arrows point to the IC, while the white arrows point to the corpus callosum.



(b) Mid-brain coronal slices of average DTI maps for 5 of our trained models (we used $\lambda_{DTI} = 2.0$, and $\lambda_{FA} = 0.1$ for the **attention** models), together with the fixed DTI atlas and FA map. Yellow arrows point to the IC, while the white arrows point to the corpus callosum.

Figure 6.6: Average DT images from 10 subjects scanned around 40 weeks PMA.

6.4 Discussion and future work

In this study, we extended the previously proposed multi-channel image registration network from Chapter 5 with layers capable of dealing with higher-order DT data. More specifically, two types of experiments were conducted. First, 2D mid-brain axial slices were used as a preliminary analysis to better understand the influence of the global weight which balances the two modalities. Here, it was found that increasing the weight in favor of the DTI channel leads to a decrease in Dice scores and an increase in average surface distances for the cGM and WM structures, while the opposite happens for the ventricles and IC. A qualitative analysis of average attention maps (see Figure 6.4) showed that a low value of the global weight ($\lambda_{DTI} < 1.0$) leads to very little influence from the microstructural channel, whereas a higher weight ($\lambda_{DTI} \geq 1.0$) leads to increasingly more pronounced DTI influence.

Weights of $\lambda_{DTI} \in [1.0, 1.5, 2.0]$ were chosen to train the 3D networks, as it was hypothesized that, given the small drop in Dice scores (~ 0.002 for cGM and ~ 0.005 for WM) and the small increase in average surface distances (~ 0.001 for cGM and ~ 0.006 for WM), it would benefit the alignment of the IC and the underlying microstructure. The 3D experiments found that for $\lambda_{DTI} \in [1.0, 1.5, 2.0]$ the IC was indeed aligned as well as the FA-only, the DTI-only, or the previously proposed attention-based T_2w+FA models (see Table 6.4). However, the drop in performance persisted in terms of cGM and WM.

On the other hand, the evaluation of white matter alignment showed that using DTI helps with aligning the underlying WM structures. This was backed by the results shown in Figure 6.3 b, where all of the models which used DTI data had significantly higher OVL values (two-sample, Wilcoxon signed rank test with 5% significance level, $p\text{-value} < 0.05$) when compared with the other models or the initial affine alignment. Moreover, the experiment showed that using **attention** when combining the T_2w and DTI data is better than not using attention, as the average OVL values in both WM and IC voxels were significantly higher ($p\text{-value} < 0.05$) when compared to the **baseline** T_2w+DTI . Finally, a qualitative analysis of average DTI maps obtained with 5 of the trained models also showed that using DTI data helps with achieving more coherent orientations of the diffusion tensors (see Figure 6.6).

The development of accurate methods for neonatal MR image registration, specifically for DTI, carries significant clinical importance. For example, DTI provides valuable insights into the microstructural properties and connectivity of white matter in the neonatal brain. Accurate image registration facilitates the alignment of DTI data across different subjects or time points, enabling the evaluation of white matter maturation and tracking developmental changes. It aids in studying the formation of white matter tracts, detecting abnormalities, and understanding the impact of early-life events or interventions on brain connectivity [379].

However, deploying models that perform neonatal MR multi-channel T_2w and DTI registration in a clinical setting has its challenges. For example, our current framework faces practical challenges due to the limited availability of diffusion data compared to structural MRI data. Nevertheless, we can consider a potential solution inspired by the work of Hu *et al.* [240] and explored by us in [55]. In this approach, instead of incorporating DTI data as input to the network, it is only used in the loss function. As a result, during inference, the trained network can successfully register pairs of T_2w images without the need to provide the extra microstructural information. This is particularly useful when higher-order data is absent in the test dataset. Furthermore, diffusion weighted MR protocols generally have a long acquisition time, during which subject motion becomes unavoidable, especially among pediatric populations. Moreover, it is frequently plagued with physiological noise, and has limited signal-to-noise ratio [385]. These challenges can adversely affect the accuracy of registration algorithms, resulting in erroneous or suboptimal outcomes.

Our study was focused on extending the previous model with higher-order DTI data, without further exploring other avenues in terms of network architecture. For example, we limited ourselves to the use of the same sized networks and the same latent space size of 32, which could potentially have a detrimental effect in our network’s capacity, as the DT data is introducing more input information. Moreover, we chose our hyperparameters based on a 2D study, which might not have a direct transfer to the more complex 3D case. This, however, can be explored in future work by looking at how different weights in the 3D models will have an effect on the output predictions. Furthermore, as described in Chapter 2.1.4, an alternative loss function for aligning DTI data is the *euclidean distance squared between deviatoric tensors*. Using \mathcal{D}_{DDS} (equation 2.28) instead of \mathcal{D}_{EDS} (equation 2.26) has the potential to further improve alignment, as \mathcal{D}_{DDS} is less sensitive to the isotropic components of diffusion tensors [386].

Similar to Chapter 5, this study did not include a comparison against more classical image registration methods, such as DTI-TK [61], and it has also relied on the assumption that the available Draw-EM tissue maps are accurate, with no opportunity to compare these results against medical raters. Finally, other higher-order diffusion data can be used in image registration applications, besides the rank-2 diffusion tensor. For example, diffusion ODF data are able to alleviate some of the limitations of the DTI model [116, 117], such as its inability to resolve crossing fibers, and have been shown to produce accurate alignment of diffusion MRI data [51]. This could be explored in future work.

Conclusions

This PhD thesis has presented the investigation and development of deep learning tools suited for analysis of multi-modal neonatal brain MRI. As a prerequisite for the contributions, the **first chapter** introduces the neonatal brain, and describes the two main MR imaging modalities utilised throughout this thesis, *i.e.*, structural T_2w MRI, and microstructural DTI, with the aim of highlighting the challenges in analysis of neonatal MRI. Moreover, **this chapter** includes information on two neonatal datasets, *i.e.*, dHCP [11] and ePrime [35], and describes their differences in terms of the cohorts, the acquisition protocols and preprocessing pipelines. The **second** and **third** chapters lay the groundwork for the methods used throughout this thesis, with a focus on classical and deep learning image registration and segmentation algorithms.

In **Chapter 4** (*Harmonised segmentation of neonatal brain MRI*) the aim was to predict tissue segmentation maps on T_2w MRI data of an unseen preterm-born neonatal population. This was achieved through investigating two unsupervised DA techniques with the objective of finding the best solution for harmonising tissue segmentation maps. Validation of data harmonisation was performed between subsamples of the dHCP and ePrime cohorts which showed comparable GA at birth and PMA at time of scan, as well as similar gender and maternal ethnicity. The evaluation found that four of the trained methods (*baseline with augmentation*, *latent with augmentation*, *image* and *image with augmentation*) achieved comparable values when compared to the dHCP subset, in terms of tissue volumes, mean global cortical thickness measures, and mean local cortical thickness measures. In fact, one hypothesis is that these four methods provided the best overall results because either they were trained using data augmentation or, in case of the image space DA method, the generator acted as a deep learning-based augmentation technique [366], which made the segmentation network more robust to the different contrast, population bias and acquisition protocol of the ePrime dataset. Nevertheless, the proposed *image with augmentation* model performed best, whereby ePrime values, tending towards higher values before harmonisation, were brought

downwards into a comparable range of values to dHCP, for 10 out of 11 cortical subregions (see Figure 4.8 last column). Moreover, a qualitative assessment showed that the proposed model corrected misclassified voxels which were prevalent in the other 3 methods (see Figure 4.9), while also outperforming the original Draw-EM segmentation by correcting a region of WM which was wrongly classified as CSF (see Figure 4.10). Finally, the harmonised cortical segmentation maps were utilised to look at differences in both global and local cortical thickness measures between term and preterm-born neonates. This analysis showed that the harmonised cortical gray matter maps resulted in global thickness measures which were comparable with the dHCP-only neonates, which was not the case before harmonisation. Moreover, significant differences between term and preterm-born neonates were found in terms of local cortical thickness measures, consistent with previous studies [367] in an adolescent cohort. Lastly, the importance of harmonising the cortical tissue maps is shown through investigating the association between neonatal cortical thickness and a language outcome measure. This analysis demonstrated that without data harmonisation, pooling images from separate datasets could lead to spurious findings that are driven by systematic differences in acquisitions rather than by true underlying effects.

In **Chapter 5** (*Attention-driven multi-channel deformable registration of structural and microstructural neonatal data*), the aim was to develop a multi-channel deep learning image registration framework capable of combining information from T_2w neonatal scans with DWI-derived FA maps. This was achieved through an attention-driven multi-channel deep learning image registration framework which selects the most salient features from the two image modalities to improve alignment of individual MR images to a common atlas space. A comparison study was performed to evaluate the results against registration networks trained on T_2w -only, FA-only, or both modalities at the same time, either with or without attention. Visual attention network blocks, such as those proposed in [309, 314], were also explored, as well as an uncertainty-aware mechanism [374] which we previously proposed. This 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.

In **Chapter 6** (*Diffusion tensor driven deep learning image registration*) the aim was to extend the previously proposed attention-based multi-channel deep learning image registration framework to deal with higher-order DT image data. The motivation for this was that registration of DT images has the potential to better align WM structures than using structural MRI only. Moreover, unlike the scalar-valued FA maps, DTI enables the inclusion of fiber orientation at each voxel. To achieve this, the networks were extended with layers which account for the change in orientation of diffusion tensors induced by the predicted deformation fields. More specifically, the *finite strain* strategy [105] was employed to reorient the tensors without changing the local microstructure. This study found that a good balance between the two modalities is harder to achieve with DTI when compared to FA maps. This could be due to the FA data having little to no contrast in the cGM ribbon, whereas DT

images are rich in information across the entire brain, making the attention network more likely to choose from both channels when aligning this structure (see also Figure 6.5). Nevertheless, the results show the importance of including fiber orientation at each voxel through the use of DTI, as the underlying microstructure is more coherent (see Figure 6.6) and better aligned with the fixed atlas (see Figure 6.3).

7.1 Limitations and future work

Each results chapter presented a summary of the main contributions, limitations and possible future directions. In this section, the aim is to highlight key limitations which can become future avenues of research.

7.1.1 Inclusion of multiple imaging modalities and labels

One possible future research avenue is to further explore the rich information present in both the dHCP and ePrime datasets. More specifically, as described in **Chapter 1.2.3**, these databases contain diffusion and functional MRI, both of which have not been explored in terms of data harmonisation.

For the deep learning image registration networks, the available Draw-EM dHCP segmentation maps could act as a guide to improve alignment of cGM and WM tissue types. Moreover, as previously proposed by [183] and explored by us in [55], the labels need not become input to the networks, but could be used only when training (in the loss function), thus not making them invaluable to prediction tasks. In fact, the inclusion of masks to help guide the registration has been previously explored by our group in Uus *et al.* [51], as well as in [148, 387], where the cortex label was used as an extra channel in order to improve the accuracy of cortical alignment which was otherwise decreased by the use of microstructural data.

7.1.2 Further exploring the image synthesis avenue

There are two potential improvements which can be brought to the proposed data harmonisation and multi-channel registration solutions. First, one improvement to the data harmonisation pipeline is the use of a Cycle-GAN architecture instead of the more simple image translation approach. In fact, in our preliminary experiments which were conducted on 2D data for prototyping the solution, we did explore this approach, but due to high memory consumption we were not able to translate it to 3D. However, as was described in the literature review (**Chapter 3.2.4**), Li *et al.* [347] or Chen *et al.* [348] proposed methods for image-level domain transfer

in multiple stages, instead of training the framework end-to-end. This could be a potential avenue to explore when extending the data harmonisation framework to diffusion MRI, or even to different cohorts, such as 6 months old infants. This way we could separately train more advanced segmentation networks, such as Attention U-Net [265], as well as use the more stable Cycle-GAN for image-to-image synthesis.

Second, the image registration network could benefit from a joint intensity and geometrical changes framework, more specifically through the use of contrast transfer, in order to separate the impact of tissue maturation on the morphological changes that happen in neonates. This is because medical image registration methods can be misguided by changes in MR contrast related to development, which reduces their sensitivity to effects related to preterm birth and early signs of disease. For this reason, one potential avenue for future work would be to account for both changes in geometry, as well as the MR intensities which locally vary throughout development due to maturation processes [51].

7.1.3 Identifying abnormal developmental patterns

Finally, the multi-modal framework developed in this thesis should be used in an application context for identifying abnormalities in the neonatal brain. For example, volumetric changes, using tensor-based morphometry (TBM), could be explored in a comparison study between term and preterm-born neonates. Moreover, as previously shown in Modat *et al.* [103], using both structural and microstructural data in a joint image registration framework can have a significant effect on the areas identified by TBM studies. It would be interesting to explore if the proposed attention-based multi-modality registration framework would have an impact in the volume change differences obtained when compared to the single-modality networks.

Appendices

A Grouping of cortical substructures 1/2

Tissue name	Cortical subregion
Insula left	Insula (left)
Insula right	Insula (right)
Occipital lobe left	Occipital (left)
Occipital lobe right	Occipital (right)
Cingulate gyrus (anterior part right)	Cingulate
Cingulate gyrus (anterior part left)	
Cingulate gyrus (posterior part right)	
Cingulate gyrus (posterior part left)	
Frontal lobe left	Frontal (left)
Frontal lobe right	Frontal (right)
Parietal lobe left	Parietal (left)
Parietal lobe right	Parietal (right)

Table 1: Grouping of cortical substructures showing their original tissue name obtained from Draw-EM [148] on the first column and their corresponding cortical subregion on the second column.

B Grouping of cortical substructures 2/2

Tissue name	Cortical subregion	
Anterior temporal lobe (medial part left)	Temporal (left)	
Anterior temporal lobe (lateral part left)		
Gyri parahippocampalis et ambiens (anterior part left)		
Superior temporal gyrus (middle part left)		
Medial and inferior temporal gyri (anterior part left)		
Lateral occipitotemporal gyrus, gyrus fusiformis (anterior part left)		
Gyri parahippocampalis et ambiens (posterior part left)		
Lateral occipitotemporal gyrus, gyrus fusiformis (posterior part left)		
Medial and inferior temporal gyri (posterior part left)		
Superior temporal gyrus (posterior part left)		
Anterior temporal lobe (medial part right)		Temporal (right)
Anterior temporal lobe (lateral part right)		
Gyri parahippocampalis et ambiens (anterior part right)		
Superior temporal gyrus (middle part right)		
Medial and inferior temporal gyri (anterior part right)		
Lateral occipitotemporal gyrus, gyrus fusiformis (anterior part right)		
Gyri parahippocampalis et ambiens (posterior part right)		
Lateral occipitotemporal gyrus, gyrus fusiformis (posterior part right)		
Medial and inferior temporal gyri (posterior part right)		
Superior temporal gyrus, posterior part right		

Table 2: Grouping of cortical substructures showing their original tissue name obtained from Draw-EM [148] on the first column and their corresponding cortical subregion on the second column. This table continues from the table above.

List of Abbreviations

$T_1\mathbf{w}$ T_1 -weighted

$T_2\mathbf{w}$ T_2 -weighted

AE autoencoder

ANN artificial neural network

ANTs advanced normalization tools

ASD average surface distance

BCE binary cross entropy

BE bending energy

CAE convolutional autoencoder

CBAM convolutional block attention module

CC cross correlation

CE cross entropy

cGM cortical gray matter

CNN convolutional neural network

CSF cerebrospinal fluid

CT computed tomography

CVAE conditional variational autoencoder

DA domain adaptation

dGM deep gray matter

DL Dice loss

Draw-EM developing region annotation with expectation maximisation

DSC	Dice score coefficient
DT	diffusion tensor
DTI	diffusion tensor imaging
DW	diffusion weighted
DW-MRI	diffusion weighted magnetic resonance imaging
DWI	diffusion weighted imaging
ELU	Exponential Linear Unit
EM	expectation maximisation
FA	fractional anisotropy
FCNN	fully convolutional neural network
FFD	free-form deformation
FLAIR	fluid attenuated inversion recovery
FN	false negatives
FNR	false negative rate
FOR	false omission rate
FP	false positives
FPR	false positive rate
GA	gestational age
GAN	generative adversarial network
GDL	generalised Dice loss
GM	gray matter
GMM	Gaussian mixture models
GPU	graphics processing unit
HARDI	high angular resolution diffusion imaging
HD	Hausdorff distance
IC	internal capsule
JE	joint entropy
K-NN	K-nearest neighbours
KL	Kullback-Leibler

LDDMM large deformations diffeomorphic metric mapping
LNCC local normalised cross correlation
LSTM long short-term memory
MD mean diffusivity
MI mutual information
MLP multilayer perceptron
MR magnetic resonance
MRI magnetic resonance imaging
MSE mean squared error
NCC normalised cross correlation
NMI normalised mutual information
NN neural network
nnU-Net no-new-Net
ODE ordinary differential equation
ODF orientation distribution functions
OVL overlap of eigenvalue-eigenvector pairs
PET positron emission tomography
PMA post-menstrual age
PPV positive predicted value
PReLU Parametric Rectified Linear Unit
ReLU Rectified Linear Unit
ROIs regions of interest
rsfMRI resting state functional MRI
SAD sum of absolute differences
SAM statistical appearance models
SE-EPI spin echo echo-planar imaging
SHARD spherical harmonics and radial decomposition
SNR signal-to-noise ratio
SPD symmetric positive-definite

SPM	statistical parametric mapping
SSD	sum of squared differences
SVD	singular value decomposition
SVF	stationary velocity field
TN	true negatives
TNR	true negative rate
TP	true positives
TPR	true positive rate
TPS	thin-plate spline
TRUS	transrectal ultrasound
TSE	turbo spin echo
US	ultrasound
VAE	variational autoencoder
WCE	weighted cross entropy
WHO	World Health Organisation
WM	white matter

Bibliography

- [1] Hannah Blencowe, Simon Cousens, Mikkel Z Oestergaard, Doris Chou, Ann-Beth Moller, Rajesh Narwal, Alma Adler, Claudia Vera Garcia, Sarah Rohde, Lale Say, et al. National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. *The Lancet*, 379(9832):2162–2172, 2012.
- [2] Li Liu, Shefali Oza, Dan Hogan, Yue Chu, Jamie Perin, Jun Zhu, Joy E Lawn, Simon Cousens, Colin Mathers, and Robert E Black. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. *The Lancet*, 388(10063):3027–3035, 2016.
- [3] John G Sled and Revital Nossin-Manor. Quantitative MRI for studying neonatal brain development. *Neuroradiology*, 55(2):97–104, 2013.
- [4] Gregory Z Tau and Bradley S Peterson. Normal development of brain circuits. *Neuropsychopharmacology*, 35(1):147–168, 2010.
- [5] Raj Ladher and Gary C Schoenwolf. Making a neural tube: Neural induction and neurulation. In *Developmental neurobiology*, pages 1–20. Springer, 2005.
- [6] Joan Stiles and Terry L Jernigan. The basics of brain development. *Neuropsychology review*, 20(4):327–348, 2010.
- [7] Mary Rutherford. MRI of the neonatal brain. *Magnetic resonance imaging of the brain in preterm infants: 24 weeks’ gestation to term*, pages 25–49, 2002.
- [8] Petra S Hüppi, Bernhard Schuknecht, Chris Boesch, Emilio Bossi, Jacques Fellinger, Christoph Fusch, and Norbert Herschkowitz. Structural and neurobehavioral delay in postnatal brain development of preterm infants. *Pediatric research*, 39(5):895–901, 1996.
- [9] Jessica Dubois, Manon Benders, Arnaud Cachia, Francois Lazeyras, R Ha-Vinh Leuchter, Stéphane V Sizonenko, Cristina Borradori-Tolsa, Jean-François Mangin, and Petra Susan Hüppi. Mapping the early cortical folding process in the preterm newborn brain. *Cerebral cortex*, 18(6):1444–1454, 2008.

- [10] Zoltán Molnár and Mary Rutherford. Brain maturation after preterm birth. *Science translational medicine*, 5(168):168ps2–168ps2, 2013.
- [11] A David Edwards, Daniel Rueckert, Stephen M Smith, Samy Abo Seada, Amir Alansary, Jennifer Almalbis, Joanna Allsop, Jesper Andersson, Tomoki Arichi, Sophie Arulkumaran, et al. The developing human connectome project neonatal data release. *Frontiers in Neuroscience*, 16, 2022.
- [12] Maria Kuklisova-Murgasova, Paul Aljabar, Latha Srinivasan, Serena J Counsell, Valentina Doria, Ahmed Serag, Ioannis S Gousias, James P Boardman, Mary A Rutherford, A David Edwards, et al. A dynamic 4d probabilistic atlas of the developing brain. *NeuroImage*, 54(4):2750–2763, 2011.
- [13] Siying Wang, Maria Kuklisova-Murgasova, Joseph V Hajnal, Christian Ledig, and Julia A Schnabel. Regression analysis for assessment of myelination status in preterm brains with magnetic resonance imaging. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 278–281. IEEE, 2016.
- [14] F. Bloch. Nuclear induction. *Physical Review*, 70(7-8):460–474, 1946.
- [15] Em Purcell, Hc Torrey, and Rv Pound. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review*, 69(1-2):37–38, 1946.
- [16] PC C Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242(5394):190–191, 1973.
- [17] Peter Mansfield and Peter K Grannell. NMR ‘diffraction’ in solids? *Journal of Physics C: solid state physics*, 6(22):L422, 1973.
- [18] Robert W Brown, Y-C Norman Cheng, E Mark Haacke, Michael R Thompson, and Ramesh Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014.
- [19] Donald W McRobbie, Elizabeth A Moore, Martin J Graves, and Martin R Prince. *MRI from Picture to Proton*. Cambridge university press, 2017.
- [20] Pankaj Gupta, Kushaljit Singh Sodhi, Akshay Kumar Saxena, Niranjan Khandelwal, and Pratibha Singhi. Neonatal cranial sonography: a concise review for clinicians. *Journal of pediatric neurosciences*, 11(1):7, 2016.
- [21] Jonathan Pindrik, Xiaobu Ye, Boram Grace Ji, Courtney Pendleton, and Edward S Ahn. Anterior fontanelle closure and size in full-term children based on head computed tomography. *Clinical pediatrics*, 53(12):1149–1157, 2014.
- [22] Monica Epelman, Alan Daneman, Christian J Kellenberger, Abdul Aziz, Osnat Konen, Rahim Moineddin, Hilary Whyte, and Susan Blaser. Neonatal encephalopathy: a prospective comparison of head US and MRI. *Pediatric radiology*, 40(10):1640–1650, 2010.
- [23] D Le Bihan. Diffusion, perfusion and functional magnetic resonance imaging. *Journal des maladies vasculaires*, 20(3):203–214, 1995.

-
- [24] John S Allen, Hanna Damasio, and Thomas J Grabowski. Normal neuroanatomical variation in the human brain: an MRI-volumetric study. *American Journal of Physical Anthropology: The Official Publication of the American Association of Physical Anthropologists*, 118(4):341–358, 2002.
- [25] Antonios Makropoulos, Paul Aljabar, Robert Wright, Britta Hüning, Nazakat Merchant, Tomoki Arichi, Nora Tusor, Joseph V Hajnal, A David Edwards, Serena J Counsell, et al. Regional growth and atlas of the developing human brain. *Neuroimage*, 125:456–478, 2016.
- [26] Peter J Basser, James Mattiello, and Denis LeBihan. MR diffusion tensor spectroscopy and imaging. *Biophysical journal*, 66(1):259–267, 1994.
- [27] Peter J Basser and Carlo Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. *Journal of magnetic resonance*, 213(2):560–570, 2011.
- [28] Pratik Mukherjee, JI Berman, Stephen W Chung, CP Hess, and RG Henry. Diffusion tensor MR imaging and fiber tractography: theoretic underpinnings. *American journal of neuroradiology*, 29(4):632–641, 2008.
- [29] Heidi Johansen-Berg and Timothy EJ Behrens. *Diffusion MRI: from quantitative measurement to in vivo neuroanatomy*. Academic Press, 2013.
- [30] Susumu Mori. *Introduction to diffusion tensor imaging*. Elsevier, 2007.
- [31] Jacques-Donald Tournier, Susumu Mori, and Alexander Leemans. Diffusion tensor imaging and beyond. *Magnetic resonance in medicine*, 65(6):1532, 2011.
- [32] Andrew L Alexander, Khader Hasan, Gordon Kindlmann, Dennis L Parker, and Jay S Tsuruda. A geometric analysis of diffusion tensor measurements of the human brain. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 44(2):283–291, 2000.
- [33] Andrew L Alexander, Samuel A Hurley, Alexey A Samsonov, Nagesh Adluru, Ameer Pasha Hosseinbor, Pouria Mossahebi, Do PM Tromp, Elizabeth Zakaszewski, and Aaron S Field. Characterization of cerebral white matter properties using quantitative magnetic resonance imaging stains. *Brain connectivity*, 1(6):423–446, 2011.
- [34] Carlo Pierpaoli and Peter J Basser. Toward a quantitative assessment of diffusion anisotropy. *Magnetic resonance in Medicine*, 36(6):893–906, 1996.
- [35] A David Edwards, Maggie E Redshaw, Nigel Kennea, Oliver Rivero-Arias, Nuria Gonzales-Cinca, Phumza Nongena, Moegamad Ederies, Shona Falconer, Andrew Chew, Omar Omar, et al. Effect of MRI on preterm infants and their families: a randomised trial with nested diagnostic and economic evaluation. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, 103(1):F15–F21, 2018.

-
- [36] Said Pertuz, Domenec Puig, and Miguel Angel Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recognition*, 46(5):1415–1432, 2013.
- [37] José Luis Pech-Pacheco, Gabriel Cristóbal, Jesús Chamorro-Martinez, and Joaquín Fernández-Valdivia. Diatom autofocusing in brightfield microscopy: a comparative study. In *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, volume 3, pages 314–317. IEEE, 2000.
- [38] Emer J. Hughes, Tobias Winchman, Francesco Padormo, Rui Teixeira, Julia Wurie, Maryanne Sharma, Matthew Fox, Jana Hutter, Lucilio Cordero-Grande, Anthony N. Price, Joanna Allsop, Jose Bueno-Conde, Nora Tusor, Tomoki Arichi, A. D. Edwards, Mary A. Rutherford, Serena J. Counsell, and Joseph V. Hajnal. A dedicated neonatal brain imaging system. *Magnetic Resonance in Medicine*, 78(2):794–804, 2017.
- [39] Lucilio Cordero-Grande, Emer J. Hughes, Jana Hutter, Anthony N. Price, and Joseph V. Hajnal. Three-dimensional motion corrected sensitivity encoding reconstruction for multi-shot multi-slice MRI: Application to neonatal brain imaging. *Magnetic Resonance in Medicine*, 79(3):1365–1376, 2018.
- [40] Maria Kuklisova-Murgasova, Gerardine Quaghebeur, Mary A Rutherford, Joseph V Hajnal, and Julia A Schnabel. Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. *Medical image analysis*, 2012.
- [41] Jana Hutter, J Donald Tournier, Anthony N Price, Lucilio Cordero-Grande, Emer J Hughes, Shaihan Malik, Johannes Steinweg, Matteo Bastiani, Stamatios N Sotiropoulos, Saad Jbabdi, et al. Time-efficient and flexible design of optimized multishell hardi diffusion. *Magnetic resonance in medicine*, 79(3):1276–1292, 2018.
- [42] Jacques-Donald Tournier, Emer Hughes, Nora Tusor, Stamatios N Sotiropoulos, Saad Jbabdi, Jesper Andersson, Daniel Rueckert, A David Edwards, and Joseph V Hajnal. Data-driven optimisation of multi-shell hardi. In *Proc ISMRM*, volume 20, 2012.
- [43] Jelle Veraart, Dmitry S Novikov, Daan Christiaens, Benjamin Ades-Aron, Jan Sijbers, and Els Fieremans. Denoising of diffusion MRI using random matrix theory. *Neuroimage*, 142:394–406, 2016.
- [44] Elias Kellner, Bibek Dhital, Valerij G Kiselev, and Marco Reisert. Gibbs-ringing artifact removal based on local subvoxel-shifts. *Magnetic resonance in medicine*, 76(5):1574–1581, 2016.
- [45] Jesper LR Andersson, Stefan Skare, and John Ashburner. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage*, 20(2):870–888, 2003.
- [46] Daan Christiaens, Lucilio Cordero-Grande, Maximilian Pietsch, Jana Hutter, Anthony N Price, Emer J Hughes, Katy Vecchiato, Maria Deprez, A David Edwards, Joseph V Hajnal, et al. Scattered slice SHARD reconstruction for

- motion correction in multi-shell diffusion MRI of the neonatal brain. *arXiv preprint arXiv:1905.02996*, 2019.
- [47] Gareth Ball, Paul Aljabar, Phumza Nongena, Nigel Kennea, Nuria Gonzalez-Cinca, Shona Falconer, Andrew TM Chew, Nicholas Harper, Julia Wurie, Mary A Rutherford, et al. Multimodal image analysis of clinical influences on preterm brain development. *Annals of Neurology*, 2017.
- [48] Madeleine L Barnett, Nora Tusor, Gareth Ball, Andrew Chew, Shona Falconer, Paul Aljabar, Jessica A Kimpton, Nigel Kennea, Mary Rutherford, A David Edwards, et al. Exploring the multiple-hit hypothesis of preterm white matter damage using diffusion MRI. *NeuroImage: Clinical*, 17:596–606, 2018.
- [49] Ralica Dimitrova, Maximilian Pietsch, Daan Christiaens, Judit Ciarrusta, Thomas Wolfers, Dafnis Batalle, Emer Hughes, Jana Hutter, Lucilio Cordero-Grande, Anthony N Price, et al. Heterogeneity in brain microstructural development following preterm birth. *Cerebral Cortex*, 30(9):4800–4810, 2020.
- [50] Mary Rutherford, Miriam Martinez Biarge, Joanna Allsop, Serena Counsell, and Frances Cowan. MRI of perinatal brain injury. *Pediatric Radiology*, 40(6):819–833, 2010.
- [51] Alena Uus, Irina Grigorescu, Maximilian Pietsch, Dafnis Batalle, Daan Christiaens, Emer Hughes, Jana Hutter, Lucilio Cordero Grande, Anthony N. Price, Jacques-Donald Tournier, Mary A. Rutherford, Serena J. Counsell, Joseph V. Hajnal, A. David Edwards, and Maria Deprez. Multi-channel 4D parametrized atlas of macro- and microstructural neonatal brain development. *Frontiers in Neuroscience*, 15:721, 2021.
- [52] Marie P Pittet, Lana Vasung, Petra S Huppi, and Laura Merlini. Newborns and preterm infants at term equivalent age: A semi-quantitative assessment of cerebral maturity. *NeuroImage: Clinical*, 24:102014, 2019.
- [53] Kenichi Oishi, Linda Chang, and Hao Huang. Baby brain atlases. *NeuroImage*, 185:865–880, 2019.
- [54] Brian Avants, Jeffrey T Duda, Hui Zhang, and James C Gee. Multivariate normalization with symmetric diffeomorphisms for multivariate studies. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 359–366. Springer, 2007.
- [55] Irina Grigorescu, Alena Uus, Daan Christiaens, Lucilio Cordero-Grande, Jana Hutter, A. David Edwards, Joseph V. Hajnal, Marc Modat, and Maria Deprez. Diffusion tensor driven image registration: A deep learning approach. In Žiga Špiclin, Jamie McClelland, Jan Kybic, and Orcun Goksel, editors, *Biomedical Image Registration*, pages 131–140, Cham, 2020. Springer International Publishing.
- [56] Alena Uus, Maximilian Pietsch, Irina Grigorescu, Daan Christiaens, Jacques-Donald Tournier, Lucilio Cordero Grande, Jana Hutter, David Edwards, Joseph Hajnal, and Maria Deprez. Multi-channel registration for diffusion

- MRI: Longitudinal analysis for the neonatal brain. In Žiga Špiclin, Jamie McClelland, Jan Kybic, and Orcun Goksel, editors, *Biomedical Image Registration*, pages 111–121, Cham, 2020. Springer International Publishing.
- [57] Daniel Forsberg, Yogesh Rathi, Sylvain Bouix, Demian Wassermann, Hans Knutsson, and Carl-Fredrik Westin. Improving registration using multi-channel diffeomorphic demons combined with certainty maps. In *International Workshop on Multimodal Brain Image Analysis*, pages 19–26. Springer, 2011.
- [58] Hidemasa Takao, Naoto Hayashi, and Kuni Ohtomo. Effect of scanner in longitudinal studies of brain volume changes. *Journal of Magnetic Resonance Imaging*, 34(2):438–444, 2011.
- [59] Russell T Shinohara, Jiwon Oh, Govind Nair, Peter A Calabresi, Christos Davatzikos, Jimit Doshi, Roland G Henry, Gloria Kim, Kristin A Linn, Nico Papinutto, et al. Volumetric analysis from a harmonized multisite brain MRI study of a single subject with multiple sclerosis. *American Journal of Neuro-radiology*, 38(8):1501–1509, 2017.
- [60] Brian B Avants, Charles L Epstein, Murray Grossman, and James C Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008.
- [61] Hui Zhang, Paul A Yushkevich, Daniel C Alexander, and James C Gee. Deformable registration of diffusion tensor MR images with explicit orientation optimization. *Medical image analysis*, 10(5):764–785, 2006.
- [62] John Ashburner. A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113, 2007.
- [63] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: A survey. *IEEE transactions on medical imaging*, 32(7):1153–1190, 2013.
- [64] G E Christensen, R D Rabbitt, and M I Miller. 3d brain mapping using a deformable neuroanatomy. *Physics in Medicine and Biology*, 39(3):609–618, mar 1994.
- [65] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. Deformable templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447, Oct 1996.
- [66] Vincent Arsigny, Olivier Commowick, Xavier Pennec, and Nicholas Ayache. A log-euclidean framework for statistics on diffeomorphisms. In Rasmus Larsen, Mads Nielsen, and Jon Sporring, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, pages 924–931, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.
- [67] J.-P. Thirion. Image matching as a diffusion process: an analogy with maxwell’s demons. *Medical Image Analysis*, 2(3):243 – 260, 1998.

-
- [68] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 1999.
- [69] Marc Modat, Gerard R Ridgway, Pankaj Daga, M Jorge Cardoso, David J Hawkes, John Ashburner, and Sébastien Ourselin. Log-euclidean free-form deformation. In *Medical Imaging 2011: Image Processing*, volume 7962, pages 541–546. SPIE, 2011.
- [70] Joseph V Hajnal and Derek LG Hill. *Medical image registration*. CRC press, 2001.
- [71] Paul Viola and William M Wells III. Alignment by maximization of mutual information. *International journal of computer vision*, 24(2):137–154, 1997.
- [72] Joseph V Hajnal, Nadeem Saeed, Elaine J Soar, Angela Oatridge, Ian R Young, and Graeme M Bydder. A registration and interpolation procedure for subvoxel matching of serially acquired MR images. *Journal of computer assisted tomography*, 19(2):289–296, 1995.
- [73] Joseph V Hajnal, Nadeem Saeed, Angela Oatridge, Elaine J Williams, Ian R Young, and Graeme M Bydder. Detection of subtle brain changes using sub-voxel registration and subtraction of serial MR images. *Journal of computer assisted tomography*, 19(5):677–691, 1995.
- [74] Karl J Friston, John Ashburner, Christopher D Frith, J-B Poline, John D Heather, and Richard SJ Frackowiak. Spatial registration and normalization of images. *Human brain mapping*, 3(3):165–189, 1995.
- [75] John Ashburner and Karl J Friston. Nonlinear spatial normalization using basis functions. *Human brain mapping*, 7(4):254–266, 1999.
- [76] J. P. Lewis. Fast template matching. *Vision Interface: Canadian Image Processing and Pattern Recognition Society*, pages 120–123, 1995.
- [77] Alexis Roche, Grégoire Malandain, Xavier Pennec, and Nicholas Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 1115–1124. Springer, 1998.
- [78] Marco Lorenzi, Nicholas Ayache, Giovanni B Frisoni, Xavier Pennec, Alzheimer’s Disease Neuroimaging Initiative (ADNI, et al. Lcc-demons: a robust and accurate symmetric diffeomorphic registration algorithm. *NeuroImage*, 81:470–483, 2013.
- [79] Pascal Cachier, Eric Bardinet, Didier Dormont, Xavier Pennec, and Nicholas Ayache. Iconic feature based nonrigid registration: the pasha algorithm. *Computer vision and image understanding*, 89(2-3):272–298, 2003.
- [80] Claude E Shannon. A mathematical theory of communication (parts i and ii). *Bell System technical journal*, pages 379–423, 1948.

-
- [81] David Mattes, David R Haynor, Hubert Vesselle, Thomas K Lewellen, and William Eubank. Pet-ct image registration in the chest using free-form deformations. *IEEE transactions on medical imaging*, 22(1):120–128, 2003.
- [82] William M Wells III, Paul Viola, Hideki Atsumi, Shin Nakajima, and Ron Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical image analysis*, 1(1):35–51, 1996.
- [83] Colin Studholme, Derek LG Hill, and David J Hawkes. Multiresolution voxel similarity measures for MR-PET registration. In *Information processing in medical imaging*, volume 3, pages 287–298. Dordrecht, The Netherlands: Kluwer, 1995.
- [84] André Collignon, Dirk Vandermeulen, Paul Suetens, and Guy Marchal. 3d multi-modality medical image registration using feature space clustering. In *International Conference on Computer Vision, Virtual Reality, and Robotics in Medicine*, pages 195–204. Springer, 1995.
- [85] Josien PW Pluim, JB Antoine Maintz, and Max A Viergever. Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging*, 22(8):986–1004, 2003.
- [86] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 16(2):187–198, 1997.
- [87] Colin Studholme, Derek LG Hill, and David J Hawkes. An overlap invariant entropy measure of 3d medical image alignment. *Pattern recognition*, 32(1):71–86, 1999.
- [88] Bernd Fischer and Jan Modersitzki. Fast diffusion registration. *Contemporary Mathematics*, 313:117–128, 2002.
- [89] Berthold KP Horn and Brian G Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981.
- [90] Grace Wahba. *Spline models for observational data*. SIAM, 1990.
- [91] Torsten Rohlfing, Calvin R Maurer, David A Bluemke, and Michael A Jacobs. Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. *IEEE transactions on medical imaging*, 22(6):730–741, 2003.
- [92] Michaël Sdika. A fast nonrigid image registration with constraints on the jacobian using large scale constrained optimization. *IEEE transactions on medical imaging*, 27(2):271–281, 2008.
- [93] J-P Thirion. Image matching as a diffusion process: an analogy with maxwell’s demons. *Medical image analysis*, 2(3):243–260, 1998.
- [94] M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157, 2005.

-
- [95] Monica Hernandez and Alejandro F Frangi. Non-parametric geodesic active regions: Method and evaluation for cerebral aneurysms segmentation in 3dra and cta. *Medical image analysis*, 11(3):224–241, 2007.
- [96] Monica Hernandez, Matias N Bossa, and Salvador Olmos. Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. *International Journal of Computer Vision*, 85(3):291–306, 2009.
- [97] Tom Vercauteren, Xavier Pennec, Aymeric Perchant, and Nicholas Ayache. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45:s61–s72, 2009.
- [98] Gary E Christensen and Hans J Johnson. Consistent image registration. *IEEE transactions on medical imaging*, 20(7):568–582, 2001.
- [99] Mirza Faisal Beg and Ali Khan. Symmetric data attachment terms for large deformation image registration. *IEEE transactions on medical imaging*, 26(9):1179–1189, 2007.
- [100] Dinggang Shen and Christos Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE transactions on medical imaging*, 21(11):1421–1439, 2002.
- [101] Daniel Rueckert, Paul Aljabar, Rolf A Heckemann, Joseph V Hajnal, and Alexander Hammers. Diffeomorphic registration using b-splines. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 702–709. Springer, 2006.
- [102] Marc Modat, Pankaj Daga, M Jorge Cardoso, Sebastien Ourselin, Gerard R Ridgway, and John Ashburner. Parametric non-rigid registration using a stationary velocity field. In *2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 145–150. IEEE, 2012.
- [103] Marc Modat, Ivor J Simpson, David M Cash, Jorge Cardoso, Philip Simon John Weston, Natalie Sarah Ryan, and Nick C Fox. Ic-p-082 multi-modal image registration: application to familial ad subjects. *Alzheimer's & Dementia*, 10:P46–P47, 2014.
- [104] Eloy Roura, Torben Schneider, Marc Modat, Pankaj Daga, Nils Muhkert, Declan Chard, Sebastien Ourselin, Xavier Lladó, and Claudia Gandini Wheeler-Kingshott. Multi-channel registration of fractional anisotropy and T1-weighted images in the presence of atrophy: application to multiple sclerosis. *Functional neurology*, 30(4):245, 2015.
- [105] Daniel C Alexander, Carlo Pierpaoli, Peter J Basser, and James C Gee. Spatial transformations of diffusion tensor magnetic resonance images. *IEEE transactions on medical imaging*, 20(11):1131–1139, 2001.
- [106] Ken Shoemake and Tom Duff. Matrix animation and polar decomposition. In *Graphics Interface*, volume 92, pages 258–264. Citeseer, 1992.

-
- [107] Hui Zhang. *Registration of diffusion tensor magnetic resonance images and its application to the quantitative analysis of human brain white matter*. University of Pennsylvania, 2007.
- [108] Daniel C Alexander, James C Gee, and Ruzena Bajcsy. Similarity measures for matching diffusion tensor images. In *BMVC*, pages 1–10, 1999.
- [109] Daniel C Alexander and James C Gee. Elastic matching of diffusion tensor images. *Computer Vision and Image Understanding*, 77(2):233–250, 2000.
- [110] Juan Ruiz-Alzola, C-F Westin, Simon K Warfield, C Alberola, S Maier, and Ron Kikinis. Nonrigid registration of 3d tensor medical data. *Medical image analysis*, 6(2):143–161, 2002.
- [111] Yan Cao, Michael I Miller, Raimond L Winslow, and Laurent Younes. Large deformation diffeomorphic metric mapping of vector fields. *IEEE transactions on medical imaging*, 24(9):1216–1230, 2005.
- [112] BT Thomas Yeo, Tom Vercauteren, Pierre Fillard, Xavier Pennec, Polina Golland, Nicholas Ayache, and Olivier Clatz. Dti registration with exact finite-strain differential. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 700–703. IEEE, 2008.
- [113] Marc Modat, Gerard R Ridgway, Zeike A Taylor, Manja Lehmann, Josephine Barnes, David J Hawkes, Nick C Fox, and Sébastien Ourselin. Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, 98(3):278–284, 2010.
- [114] James C Gee and Daniel C Alexander. Diffusion-tensor image registration. In *Visualization and processing of tensor fields*, pages 327–342. Springer, 2006.
- [115] Yi Wang, Yu Shen, Dongyang Liu, Guoqin Li, Zhe Guo, Yangyu Fan, and Yilong Niu. Evaluations of diffusion tensor image registration based on fiber tractography. *Biomedical engineering online*, 16(1):1–20, 2017.
- [116] Mette R Wiegell, Henrik BW Larsson, and Van J Wedeen. Fiber crossing in human brain depicted with diffusion tensor MR imaging. *Radiology*, 217(3):897–903, 2000.
- [117] David S Tuch, Timothy G Reese, Mette R Wiegell, Nikos Makris, John W Belliveau, and Van J Wedeen. High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 48(4):577–582, 2002.
- [118] J-Donald Tournier, Fernando Calamante, David G Gadian, and Alan Connelly. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage*, 23(3):1176–1185, 2004.
- [119] David Raffelt, J-Donald Tournier, Jurgen Fripp, Stuart Crozier, Alan Connelly, and Olivier Salvado. Symmetric diffeomorphic registration of fibre orientation distributions. *Neuroimage*, 56(3):1171–1180, 2011.

-
- [120] Adam W Anderson. Measurement of fiber orientation distributions using high angular resolution diffusion imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 54(5):1194–1206, 2005.
- [121] Saleha Masood, Muhammad Sharif, Afifa Masood, Mussarat Yasmin, and Mudassar Raza. A survey on medical image segmentation. *Current Medical Imaging*, 11(1):3–14, 2015.
- [122] Neeraj Sharma and Lalit M Aggarwal. Automated medical image segmentation techniques. *Journal of medical physics/Association of Medical Physicists of India*, 35(1):3, 2010.
- [123] Paul A Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C Gee, and Guido Gerig. User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage*, 31(3):1116–1128, 2006.
- [124] Xiao Han, Chenyang Xu, and Jerry L. Prince. A topology preserving level set method for geometric deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):755–768, 2003.
- [125] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685, 2001.
- [126] Timothy F Cootes, Andrew Hill, Christopher J Taylor, and Jane Haslam. Use of active shape models for locating structures in medical images. *Image and vision computing*, 12(6):355–365, 1994.
- [127] D Louis Collins, Colin J Holmes, Terrence M Peters, and Alan C Evans. Automatic 3d model-based neuroanatomical segmentation. *Human brain mapping*, 3(3):190–208, 1995.
- [128] Torsten Rohlfing, Robert Brandt, Randolph Menzel, and Calvin R Maurer Jr. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage*, 21(4):1428–1442, 2004.
- [129] Rolf A Heckemann, Joseph V Hajnal, Paul Aljabar, Daniel Rueckert, and Alexander Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115–126, 2006.
- [130] Xabier Artaechevarria, Arrate Munoz-Barrutia, and Carlos Ortiz-de Solorzano. Combination strategies in multi-atlas image segmentation: application to brain MR data. *IEEE transactions on medical imaging*, 28(8):1266–1277, 2009.
- [131] Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Joseph V Hajnal, and Daniel Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738, 2009.

-
- [132] Juan Eugenio Iglesias and Mert R Sabuncu. Multi-atlas segmentation of biomedical images: a survey. *Medical image analysis*, 24(1):205–219, 2015.
- [133] Robert Wright, Vanessa Kyriakopoulou, Christian Ledig, Mary A Rutherford, Joseph V Hajnal, Daniel Rueckert, and Paul Aljabar. Automatic quantification of normal cortical folding patterns from fetal brain MRI. *Neuroimage*, 91:21–32, 2014.
- [134] Hongzhi Wang and Paul A Yushkevich. Multi-atlas segmentation with joint label fusion and corrective learning an open source implementation. *Frontiers in neuroinformatics*, 7:27, 2013.
- [135] M Jorge Cardoso, Matthew J Clarkson, Gerard R Ridgway, Marc Modat, Nick C Fox, Sebastien Ourselin, Alzheimer’s Disease Neuroimaging Initiative, et al. Load: a locally adaptive cortical segmentation algorithm. *NeuroImage*, 56(3):1386–1397, 2011.
- [136] M Jorge Cardoso, Andrew Melbourne, Giles S Kendall, Marc Modat, Cornelia F Hagemann, Nicola J Robertson, Neil Marlow, and Sebastien Ourselin. Adaptive neonate brain segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 378–386. Springer, 2011.
- [137] M Jorge Cardoso, Andrew Melbourne, Giles S Kendall, Marc Modat, Nicola J Robertson, Neil Marlow, and Sebastien Ourselin. AdaPT: an adaptive preterm segmentation algorithm for neonatal brain MRI. *NeuroImage*, 65:97–108, 2013.
- [138] Stephen M Smith. Fast robust automated brain extraction. *Human brain mapping*, 2002.
- [139] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith. FSL. *Neuroimage*, 62(2):782–790, 2012.
- [140] Jimit Doshi, Guray Erus, Yangming Ou, Bilwaj Gaonkar, and Christos Davatzikos. Multi-atlas skull-stripping. *Academic radiology*, 20(12):1566–1576, 2013.
- [141] Linmin Pei, Murat Ak, Nourel Hoda M Tahon, Serafettin Zenkin, Safa Alkarawi, Abdallah Kamal, Mahir Yilmaz, Lingling Chen, Mehmet Er, Nursima Ak, et al. A general skull stripping of multiparametric brain MRIs using 3D convolutional neural network. *Scientific Reports*, 12(1):1–11, 2022.
- [142] Andrew Simmons, Paul S Tofts, Gareth J Barker, and Simon R Arridge. Sources of intensity nonuniformity in spin echo images at 1.5 t. *Magnetic resonance in medicine*, 32(1):121–128, 1994.
- [143] John G Sled, Alex P Zijdenbos, and Alan C Evans. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*, 17(1):87–97, 1998.
- [144] Nicholas J Tustison, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee. N4itk: improved n3 bias correction. *IEEE transactions on medical imaging*, 29(6):1310–1320, 2010.

-
- [145] John Ashburner and Karl J Friston. Unified segmentation. *Neuroimage*, 26(3):839–851, 2005.
- [146] Maxim Zaitsev, Julian Maclaren, and Michael Herbst. Motion artifacts in MRI: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, 42(4):887–901, 2015.
- [147] Marcel Prastawa, John H Gilmore, Weili Lin, and Guido Gerig. Automatic segmentation of MR images of the developing newborn brain. *Medical image analysis*, 9(5):457–466, 2005.
- [148] Antonios Makropoulos, Emma C Robinson, Andreas Schuh, Robert Wright, Sean Fitzgibbon, Jelena Bozek, Serena J Counsell, Johannes Steinweg, Katy Vecchiato, Jonathan Passerat-Palmbach, et al. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *Neuroimage*, 2018.
- [149] Lucilio Cordero-Grande, Rui Pedro AG Teixeira, Emer J Hughes, Jana Hut-ter, Anthony N Price, and Joseph V Hajnal. Sensitivity encoding for aligned multishot magnetic resonance reconstruction. *IEEE Transactions on Computational Imaging*, 2(3):266–280, 2016.
- [150] Simon K Warfield, Michael Kaus, Ferenc A Jolesz, and Ron Kikinis. Adaptive, template moderated, spatially varying statistical classification. *Medical image analysis*, 4(1):43–55, 2000.
- [151] Chang Wen Chen, Jiebo Luo, and Kevin J Parker. Image segmentation via adaptive k-mean clustering and knowledge-based morphological operations with biomedical applications. *IEEE transactions on image processing*, 7(12):1673–1683, 1998.
- [152] Regina Pohle and Klaus D Toennies. Segmentation of medical images using adaptive region growing. In *Medical Imaging 2001: Image Processing*, volume 4322, pages 1337–1346. SPIE, 2001.
- [153] SR Kannan, R Devi, S Ramathilagam, and K Takezawa. Effective FCM noise clustering algorithms in medical images. *Computers in biology and medicine*, 43(2):73–83, 2013.
- [154] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22, 1977.
- [155] William M Wells, W Eric L Grimson, Ron Kikinis, and Ferenc A Jolesz. Adaptive segmentation of MRI data. *IEEE transactions on medical imaging*, 15(4):429–442, 1996.
- [156] Niranjan Joshi and Michael Brady. A non-parametric model for partial volume segmentation of MR images. In *BMVC*, 2005.
- [157] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. A unifying framework for partial volume segmentation of brain MR images. *IEEE transactions on medical imaging*, 22(1):105–119, 2003.

- [158] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Automated model-based bias field correction of MR images of the brain. *IEEE transactions on medical imaging*, 18(10):885–896, 1999.
- [159] Yongyue Zhang, Michael Brady, and Stephen Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE transactions on medical imaging*, 20(1):45–57, 2001.
- [160] Koen Van Leemput, Frederik Maes, Dirk Vandermeulen, and Paul Suetens. Automated model-based tissue classification of MR images of the brain. *IEEE transactions on medical imaging*, 18(10):897–908, 1999.
- [161] Piotr A Habas, Kio Kim, Francois Rousseau, Orit A Glenn, A James Barkovich, and Colin Studholme. Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses. *Human brain mapping*, 31(9):1348–1358, 2010.
- [162] Andrew Melbourne, M Jorge Cardoso, Giles S Kendall, Nicola J Robertson, Neil Marlow, and Sebastien Ourselin. NeoBrainS12 challenge: Adaptive neonatal MRI brain segmentation with myelinated white matter class and automated extraction of ventricles I-IV. *Proceedings of the MICCAI Grand Challenge: Neonatal Brain Segmentation*, pages 16–21, 2012.
- [163] Antonios Makropoulos, Serena J Counsell, and Daniel Rueckert. A review on automatic fetal and neonatal brain MRI segmentation. *NeuroImage*, 170:231–248, 2018.
- [164] Ivana Išgum, Manon JNL Benders, Brian Avants, M Jorge Cardoso, Serena J Counsell, Elda Fisci Gomez, Laura Gui, Petra S Hüppi, Karina J Kersbergen, Antonios Makropoulos, et al. Evaluation of automatic neonatal brain segmentation algorithms: the neobrain12 challenge. *Medical image analysis*, 20(1):135–151, 2015.
- [165] Siying Wang, Maria Kuklisova-Murgasova, and Julia A Schnabel. An atlas-based method for neonatal MR brain tissue segmentation. *Proceedings of the MICCAI Grand Challenge: Neonatal Brain Segmentation*, pages 28–35, 2012.
- [166] Brian B Avants, Nicholas J Tustison, Jue Wu, Philip A Cook, and James C Gee. An open source multivariate framework for n-tissue segmentation with evaluation on public data. *Neuroinformatics*, 9(4):381–400, 2011.
- [167] Jue Wu and Brian Avants. Automatic registration-based segmentation for neonatal brains using ants and atropos. *MICCAI Grand Challenge on Neonatal Brain Segmentation*, 2012:36–47, 2012.
- [168] Navid Shiee, Pierre-Louis Bazin, Jennifer L Cuzzocreo, Ari Blitz, and Dzung L Pham. Segmentation of brain images using adaptive atlases with application to ventriculomegaly. In *Biennial International Conference on Information Processing in Medical Imaging*, pages 1–12. Springer, 2011.

- [169] Antonios Makropoulos, Christian Ledig, Paul Aljabar, Ahmed Serag, Joseph V Hajnal, A David Edwards, Serena J Counsell, and Daniel Rueckert. Automatic tissue and structural segmentation of neonatal brain MRI using expectation-maximization. *MICCAI Grand Challenge on Neonatal Brain Segmentation*, 2012:9–15, 2012.
- [170] Antonios Makropoulos, Ioannis S Gousias, Christian Ledig, Paul Aljabar, Ahmed Serag, Joseph V Hajnal, A David Edwards, Serena J Counsell, and Daniel Rueckert. Automatic whole brain MRI segmentation of the developing neonatal brain. *IEEE transactions on medical imaging*, 2014.
- [171] Ioannis S Gousias, A David Edwards, Mary A Rutherford, Serena J Counsell, Jo V Hajnal, Daniel Rueckert, and Alexander Hammers. Magnetic resonance imaging of the newborn brain: manual segmentation of labelled atlases in term-born and preterm infants. *Neuroimage*, 62(3):1499–1509, 2012.
- [172] Hui Xue, Latha Srinivasan, Shuzhou Jiang, Mary Rutherford, A David Edwards, Daniel Rueckert, and Joseph V Hajnal. Automatic segmentation and reconstruction of the cortex from neonatal MRI. *Neuroimage*, 38(3):461–477, 2007.
- [173] Abdel Aziz Taha and Allan Hanbury. Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool. *BMC medical imaging*, 15(1):1–28, 2015.
- [174] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.
- [175] Daniel P Huttenlocher, Gregory A. Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863, 1993.
- [176] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.
- [177] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.
- [178] Wenlu Zhang, Rongjian Li, Houtao Deng, Li Wang, Weili Lin, Shuiwang Ji, and Dinggang Shen. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *NeuroImage*, 108:214–224, 2015.
- [179] Holger R Roth, Christopher T Lee, Hoo-Chang Shin, Ari Seff, Lauren Kim, Jianhua Yao, Le Lu, and Ronald M Summers. Anatomy-specific classification of medical images using deep convolutional nets. *arXiv preprint arXiv:1504.04003*, 2015.
- [180] Pim Moeskops, Jelmer M Wolterink, Bas HM van der Velden, Kenneth GA Gilhuijs, Tim Leiner, Max A Viergever, and Ivana Išgum. Deep learning for multi-task medical image segmentation in multiple modalities. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 478–486. Springer, 2016.

-
- [181] Adrian V. Dalca, Guha Balakrishnan, John Guttag, and Mert R. Sabuncu. Unsupervised learning for fast probabilistic diffeomorphic registration. *Lecture Notes in Computer Science*, page 729738, 2018.
- [182] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):17881800, Aug 2019.
- [183] Yipeng Hu, Marc Modat, Eli Gibson, Wenqi Li, Nooshin Ghavami, Ester Bonmati, Guotai Wang, Steven Bandula, Caroline M. Moore, Mark Emberton, and et al. Weakly-supervised convolutional neural networks for multimodal image registration. *Medical Image Analysis*, 49:113, Oct 2018.
- [184] Heung-Il Suk, Seong-Whan Lee, Dinggang Shen, Alzheimer’s Disease Neuroimaging Initiative, et al. Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis. *NeuroImage*, 101:569–582, 2014.
- [185] Heung-Il Suk, Dinggang Shen, Alzheimers Disease Neuroimaging Initiative, et al. Deep learning in diagnosis of brain disorders. In *Recent Progress in Brain and Cognitive Engineering*, pages 203–213. Springer, 2015.
- [186] Qi Dou, Hao Chen, Lequan Yu, Lei Zhao, Jing Qin, Defeng Wang, Vincent CT Mok, Lin Shi, and Pheng-Ann Heng. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. *IEEE transactions on medical imaging*, 35(5):1182–1195, 2016.
- [187] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.
- [188] Aurélien Géron. *Hands-on machine learning with Scikit-Learn and Tensor-Flow: concepts, tools, and techniques to build intelligent systems*. " O’Reilly Media, Inc.", 2017.
- [189] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.
- [190] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [191] Sagar Sharma, Simone Sharma, and Anidhya Athaiya. Activation functions in neural networks. *Towards Data Sci*, 6(12):310–316, 2017.
- [192] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.
- [193] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.

-
- [194] David Bertoin, Jérôme Bolte, Sébastien Gerchinovitz, and Edouard Pauwels. Numerical influence of $\text{relu}(0)$ on backpropagation. *Advances in Neural Information Processing Systems*, 34:468–479, 2021.
- [195] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.
- [196] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, 2015.
- [197] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016.
- [198] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.
- [199] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.
- [200] Yuxin Wu and Kaiming He. Group normalization. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.
- [201] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [202] Leslie N. Smith. No more pesky learning rate guessing games. *CoRR*, 2015.
- [203] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.
- [204] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.
- [205] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention MICCAI 2015*, page 234241, 2015.
- [206] Fernando Pérez-García, Rachel Sparks, and Sebastien Ourselin. TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *arXiv:2003.04696 [cs, eess, stat]*, March 2020. arXiv: 2003.04696.
- [207] Benjamin Billot, Eleanor Robinson, Adrian V Dalca, and Juan Eugenio Iglesias. Partial volume segmentation of brain MRI scans of any resolution and contrast. In *International Conference on Medical image computing and computer-assisted intervention*, pages 177–187. Springer, 2020.

-
- [208] Richard Shaw, Carole Sudre, Sebastien Ourselin, and M Jorge Cardoso. MRI k-space motion artefact augmentation: model robustness and task-specific uncertainty. In *International Conference on Medical Imaging with Deep Learning—Full Paper Track*, 2018.
- [209] Carole H Sudre, M Jorge Cardoso, Sebastien Ourselin, Alzheimers Disease Neuroimaging Initiative, et al. Longitudinal segmentation of age-related white matter hyperintensities. *Medical Image Analysis*, 38:50–64, 2017.
- [210] Reuben Dorent, Thomas Booth, Wenqi Li, Carole H Sudre, Sina Kafiabadi, Jorge Cardoso, Sebastien Ourselin, and Tom Vercauteren. Learning joint segmentation of tissues and brain lesions from task-specific hetero-modal domain-shifted datasets. *Medical image analysis*, 67:101862, 2021.
- [211] Ke Yan, Jinzheng Cai, Dakai Jin, Shun Miao, Dazhou Guo, Adam P Harrison, Youbao Tang, Jing Xiao, Jingjing Lu, and Le Lu. Sam: Self-supervised learning of pixel-wise anatomical embeddings in radiological images. *IEEE Transactions on Medical Imaging*, 2022.
- [212] Asifullah Khan, Anabia Sohail, Umme Zahoor, and Aqsa Saeed Qureshi. A survey of the recent architectures of deep convolutional neural networks. *arXiv preprint arXiv:1901.06032*, 2019.
- [213] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [214] John Duchi. Derivations for linear algebra and optimization. *Berkeley, California*, 3(1):2325–5870, 2007.
- [215] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pages 424–432. Springer, 2016.
- [216] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.
- [217] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- [218] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58:101552, 2019.
- [219] Joseph Paul Cohen, Margaux Luck, and Sina Honari. Distribution matching losses can hallucinate features in medical image translation. In *International conference on medical image computing and computer-assisted intervention*, pages 529–536. Springer, 2018.
- [220] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

-
- [221] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
- [222] Taeksoo Kim, Moon-su Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. Learning to discover cross-domain relations with generative adversarial networks. In *International conference on machine learning*, pages 1857–1865. PMLR, 2017.
- [223] Jelmer M Wolterink, Anna M Dinkla, Mark HF Savenije, Peter R Seevinck, Cornelis AT van den Berg, and Ivana Išgum. Deep MR to CT synthesis using unpaired data. In *International workshop on simulation and synthesis in medical imaging*, pages 14–23. Springer, 2017.
- [224] Zizhao Zhang, Lin Yang, and Yefeng Zheng. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern Recognition*, pages 9242–9251, 2018.
- [225] Faisal Mahmood, Richard Chen, and Nicholas J Durr. Unsupervised reverse domain adaptation for synthetic medical images via adversarial training. *IEEE transactions on medical imaging*, 37(12):2572–2581, 2018.
- [226] Aïcha BenTaieb and Ghassan Hamarneh. Adversarial stain transfer for histopathology image analysis. *IEEE transactions on medical imaging*, 37(3):792–802, 2017.
- [227] Grant Haskins, Uwe Kruger, and Pingkun Yan. Deep learning in medical image registration: A survey, 2019.
- [228] Guorong Wu, Minjeong Kim, Qian Wang, Yaozong Gao, Shu Liao, and Dinggang Shen. Unsupervised deep feature learning for deformable registration of MR brain images. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8150 LNCS of *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pages 649–656, 10 2013.
- [229] Dinggang Shen. Image registration by local histogram matching. *Pattern Recognition*, 40(4):1161–1172, 2007.
- [230] Koen A. J. Eppenhof and Josien P. W. Pluim. Error estimation of deformable image registration of pulmonary CT scans using convolutional neural networks. *Journal of Medical Imaging*, 5(2):1 – 11, 2018.
- [231] Martin Simonovsky, Benjamin Gutierrez-Becker, Diana Mateus, Nassir Navab, and Nikos Komodakis. A deep metric for multimodal registration. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pages 10–18, 2016.

- [232] Robert Wright, Bishesh Khanal, Alberto Gómez, Emily Skelton, Jacqueline Matthew, Joseph V. Hajnal, Daniel Rueckert, and Julia A. Schnabel. LSTM spatial co-transformer networks for registration of 3D fetal US and MR brain images. In *DATRA/PIPPi@MICCAI*, 2018.
- [233] Mattias Paul Heinrich, Mark Jenkinson, Bartłomiej W. Papież, Sir Michael Brady, and Julia A. Schnabel. Towards realtime multimodal fusion for image-guided interventions using self-similarities. In Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pages 187–194, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.
- [234] Xiao Yang, Roland Kwitt, and Marc Niethammer. Fast predictive image registration. *Lecture Notes in Computer Science*, page 4857, 2016.
- [235] Marc-Michel Rohé, Manasi Datar, Tobias Heimann, Maxime Sermesant, and Xavier Pennec. Svf-net: Learning deformable image registration using shape matching. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 266–274. Springer, 2017.
- [236] Xiaohuan Cao, Jianhua Yang, Jun Zhang, Dong Nie, Minjeong Kim, Qian Wang, and Dinggang Shen. Deformable image registration based on similarity-steered cnn regression. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 300–308. Springer, 2017.
- [237] Hristina Uzunova, Matthias Wilms, Heinz Handels, and Jan Ehrhardt. Training cnns for image registration from few samples with model-based data augmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 223–231. Springer, 2017.
- [238] Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Hausser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. FlowNet: Learning optical flow with convolutional networks. *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015.
- [239] Jingfan Fan, Xiaohuan Cao, Pew-Thian Yap, and Dinggang Shen. Birnet: Brain image registration using dual-supervised fully convolutional networks. *Medical Image Analysis*, 54:193206, May 2019.
- [240] Yipeng Hu, Marc Modat, Eli Gibson, Nooshin Ghavami, Ester Bonmati, Caroline M. Moore, Mark Emberton, J. Alison Noble, Dean C. Barratt, and Tom Vercauteren. Label-driven weakly-supervised learning for multimodal deformable image registration. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Apr 2018.
- [241] Yipeng Hu, Eli Gibson, Nooshin Ghavami, Ester Bonmati, Caroline M. Moore, Mark Emberton, Tom Vercauteren, J. Alison Noble, and Dean C. Barratt. Adversarial deformation regularization for training image registration neural networks. *Lecture Notes in Computer Science*, page 774782, 2018.

-
- [242] Alessa Hering, Sven Kuckertz, Stefan Heldmann, and Mattias P. Heinrich. Enhancing label-driven deep deformable image registration with local distance metrics for state-of-the-art cardiac motion tracking. *Bildverarbeitung für die Medizin 2019*, page 309314, 2019.
- [243] Hongming Li and Yong Fan. Non-rigid image registration using fully convolutional networks with deep self-supervision, 2017.
- [244] Hongming Li and Yong Fan. Non-rigid image registration using self-supervised fully convolutional networks without training data, 2018.
- [245] Brian B Avants, Nicholas J Tustison, Gang Song, Philip A Cook, Arno Klein, and James C Gee. A reproducible evaluation of ants similarity metric performance in brain image registration. *Neuroimage*, 54(3):2033–2044, 2011.
- [246] Bob D. de Vos, Floris F. Berendsen, Max A. Viergever, Marius Staring, and Ivana Išgum. End-to-end unsupervised deformable image registration with a convolutional neural network. *Lecture Notes in Computer Science*, page 204212, 2017.
- [247] Bob D. de Vos, Floris F. Berendsen, Max A. Viergever, Hessam Sokooti, Marius Staring, and Ivana Išgum. A deep learning framework for unsupervised affine and deformable image registration, 2018.
- [248] Stefan Klein, Marius Staring, Keelin Murphy, Max A Viergever, and Josien PW Pluim. Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29(1):196–205, 2009.
- [249] Stergios Christodoulidis, Mihir Sahasrabudhe, Maria Vakalopoulou, Guillaume Chassagnon, Marie-Pierre Revel, Stavroula Mougiakakou, and Nikos Paragios. Linear and deformable image registration with 3d convolutional neural networks, 2018.
- [250] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, Adrian V. Dalca, and John Guttag. An unsupervised learning model for deformable medical image registration. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun 2018.
- [251] Adrian V Dalca, Guha Balakrishnan, John Guttag, and Mert R Sabuncu. Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. *Medical image analysis*, 57:226–236, 2019.
- [252] Julian Krebs, Tommaso Mansi, Boris Mailhé, Nicholas Ayache, and Hervé Delingette. Unsupervised probabilistic deformation modeling for robust diffeomorphic registration, 2018.
- [253] Julian Krebs, Hervé Delingette, Boris Mailhé, Nicholas Ayache, and Tommaso Mansi. Learning a probabilistic model for diffeomorphic registration. *IEEE transactions on medical imaging*, 38(9):2165–2176, 2019.
- [254] Dongyang Kuang and Tanya Schmah. Faim – a convnet method for unsupervised 3d medical image registration, 2018.

- [255] Jun Zhang. Inverse-consistent deep networks for unsupervised deformable image registration, 2018.
- [256] Jingfan Fan, Xiaohuan Cao, Zhong Xue, Pew-Thian Yap, and Dinggang Shen. Adversarial similarity network for evaluating image alignment in deep learning based registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 739–746. Springer, 2018.
- [257] Tony CW Mok and Albert Chung. Large deformation diffeomorphic image registration with laplacian pyramid networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 211–221. Springer, 2020.
- [258] Tycho FA van der Ouderaa, Ivana Išgum, Wouter B Veldhuis, and Bob D de Vos. Deep group-wise variational diffeomorphic image registration. In *International Workshop on Thoracic Image Analysis*, pages 155–164. Springer, 2020.
- [259] Dongdong Gu, Xiaohuan Cao, Shanshan Ma, Lei Chen, Guocai Liu, Dinggang Shen, and Zhong Xue. Pair-wise and group-wise deformation consistency in deep registration network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 171–180. Springer, 2020.
- [260] Inwan Yoo, David G. C. Hildebrand, Willie F. Tobin, Wei-Chung Allen Lee, and Won-Ki Jeong. ssemnet: Serial-section electron microscopy image registration using a spatial transformer network with learned features. *Lecture Notes in Computer Science*, page 249257, 2017.
- [261] Matthew CH Lee, Ozan Oktay, Andreas Schuh, Michiel Schaap, and Ben Glocker. Image-and-spatial transformer networks for structure-guided image registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 337–345. Springer, 2019.
- [262] Risheng Wang, Tao Lei, Ruixia Cui, Bingtao Zhang, Hongying Meng, and Asoke K Nandi. Medical image segmentation using deep learning: A survey. *IET Image Processing*, 16(5):1243–1267, 2022.
- [263] Patrick Ferdinand Christ, Florian Ettliger, Felix Grün, Mohamed Ezzeldin A Elshaera, Jana Lipkova, Sebastian Schlecht, Freba Ahmaddy, Sunil Tataavarty, Marc Bickel, Patrick Bilic, et al. Automatic liver and tumor segmentation of ct and MRI volumes using cascaded fully convolutional neural networks. *arXiv preprint arXiv:1702.05970*, 2017.
- [264] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. UNet++: A nested U-Net architecture for medical image segmentation. In *Deep learning in medical image analysis and multimodal learning for clinical decision support*, pages 3–11. Springer, 2018.
- [265] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention U-Net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

- [266] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. UNETR: Transformers for 3D medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 574–584, 2022.
- [267] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In *International MICCAI Brainlesion Workshop*, pages 272–284. Springer, 2022.
- [268] Alena U Uus, Mohammad-Usamah Ayub, Abi Gartner, Vanessa Kyriakopoulou, Maximilian Pietsch, Irina Grigorescu, Daan Christiaens, Jana Hutten, Lucilio Cordero Grande, Anthony Price, et al. Segmentation of periventricular white matter in neonatal brain MRI: Analysis of brain maturation in term and preterm cohorts. In *International Workshop on Preterm, Perinatal and Paediatric Image Analysis*, pages 94–104. Springer, 2022.
- [269] Wenqi Li, Guotai Wang, Lucas Fidon, Sebastien Ourselin, M Jorge Cardoso, and Tom Vercauteren. On the compactness, efficiency, and representation of 3d convolutional networks: brain parcellation as a pretext task. In *International conference on information processing in medical imaging*, pages 348–360. Springer, 2017.
- [270] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.
- [271] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.
- [272] Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage*, 170:456–470, 2018.
- [273] Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular psychiatry*, 19(6):659–667, 2014.
- [274] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [275] Andreas Veit, Michael J Wilber, and Serge Belongie. Residual networks behave like ensembles of relatively shallow networks. *Advances in neural information processing systems*, 29, 2016.

- [276] Hao Chen, Qi Dou, Lequan Yu, and Pheng-Ann Heng. Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation. *arXiv preprint arXiv:1608.05895*, 2016.
- [277] Xiao Xiao, Shen Lian, Zhiming Luo, and Shaozi Li. Weighted res-UNet for high-quality retina vessel segmentation. In *2018 9th international conference on information technology in medicine and education (ITME)*, pages 327–331. IEEE, 2018.
- [278] Ana Lourenço, Eric Kerfoot, Connor Dibblin, Ebrahim Alskaf, Mustafa Anjari, Anil A Bharath, Andrew P King, Henry Chubb, Teresa M Correia, and Marta Varela. Left atrial ejection fraction estimation using SEGANet for fully automated segmentation of CINE MRI. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 137–145. Springer, 2020.
- [279] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.
- [280] Steven Guan, Amir A Khan, Siddhartha Sikdar, and Parag V Chitnis. Fully dense UNet for 2-d sparse photoacoustic tomography artifact removal. *IEEE journal of biomedical and health informatics*, 24(2):568–576, 2019.
- [281] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, 2019.
- [282] Jose Dolz, Karthik Gopinath, Jing Yuan, Herve Lombaert, Christian Desrosiers, and Ismail Ben Ayed. Hyperdense-net: a hyper-densely connected cnn for multi-modal image segmentation. *IEEE transactions on medical imaging*, 38(5):1116–1126, 2018.
- [283] Li Wang, Dong Nie, Guannan Li, Élodie Puybareau, Jose Dolz, Qian Zhang, Fan Wang, Jing Xia, Zhengwang Wu, Jia-Wei Chen, et al. Benchmark on automatic six-month-old infant brain segmentation algorithms: the iseg-2017 challenge. *IEEE transactions on medical imaging*, 38(9):2219–2230, 2019.
- [284] Yang Ding, Rolando Acosta, Vicente Enguix, Sabrina Suffren, Janosch Ortman, David Luck, Jose Dolz, and Gregory A Lodygensky. Using deep convolutional neural networks for neonatal brain image segmentation. *Frontiers in neuroscience*, 14:207, 2020.
- [285] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [286] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: Context encoder

- network for 2d medical image segmentation. *IEEE transactions on medical imaging*, 38(10):2281–2292, 2019.
- [287] Saqib Qamar, Hai Jin, Ran Zheng, Parvez Ahmad, and Mohd Usama. A variant form of 3d-UNet for infant brain segmentation. *Future Generation Computer Systems*, 108:613–623, 2020.
- [288] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [289] Pim Moeskops, Max A Viergever, Adriënne M Mendrik, Linda S De Vries, Manon JNL Benders, and Ivana Išgum. Automatic segmentation of MR brain images with a convolutional neural network. *IEEE transactions on medical imaging*, 35(5):1252–1261, 2016.
- [290] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M. Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. *Lecture Notes in Computer Science*, 2017.
- [291] Paula Ramirez Gilliland, Alena Uus, Milou PM van Poppel, Irina Grigorescu, Johannes K Steinweg, David FA Lloyd, Kuberan Pushparajah, Andrew P King, and Maria Deprez. Automated multi-class fetal cardiac vessel segmentation in aortic arch anomalies using t2-weighted 3d fetal MRI. In *International Workshop on Preterm, Perinatal and Paediatric Image Analysis*, pages 82–93. Springer, 2022.
- [292] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky loss function for image segmentation using 3d fully convolutional deep networks. In *International workshop on machine learning in medical imaging*, pages 379–387. Springer, 2017.
- [293] Amos Tversky. Features of similarity. *Psychological review*, 84(4):327, 1977.
- [294] Lucas Fidon, Wenqi Li, Luis C Garcia-Peraza-Herrera, Jinendra Ekanayake, Neil Kitchen, Sébastien Ourselin, and Tom Vercauteren. Generalised wasserstein dice score for imbalanced multi-class segmentation using holistic convolutional networks. In *International MICCAI brainlesion workshop*, pages 64–76. Springer, 2017.
- [295] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.
- [296] Francesco Caliva, Claudia Iriondo, Alejandro Morales Martinez, Sharmila Majumdar, and Valentina Pedoia. Distance map loss penalty term for semantic segmentation. *arXiv preprint arXiv:1908.03679*, 2019.
- [297] Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Mattias Heinrich, Wenjia Bai, Jose Caballero, Stuart A Cook, Antonio De Marvao, Timothy Dawes, Declan P ORegan, et al. Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation. *IEEE transactions on medical imaging*, 37(2):384–395, 2017.

-
- [298] Gianni Brauwers and Flavius Frasincar. A general survey on attention mechanisms in deep learning. *IEEE Transactions on Knowledge and Data Engineering*, 2021.
- [299] Zhaoyang Niu, Guoqiang Zhong, and Hui Yu. A review on the attention mechanism of deep learning. *Neurocomputing*, 452:48–62, 2021.
- [300] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [301] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *stat*, 1050:20, 2017.
- [302] Artem Komarichev, Zichun Zhong, and Jing Hua. A-cnn: Annularly convolutional neural networks on point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7421–7430, 2019.
- [303] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.
- [304] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- [305] Amitojdeep Singh, Sourya Sengupta, and Vasudevan Lakshminarayanan. Explainable deep learning models in medical image analysis. *Journal of Imaging*, 6(6):52, 2020.
- [306] Wojciech Samek, Thomas Wiegand, and Klaus-Robert Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*, 2017.
- [307] Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pages 80–89. IEEE, 2018.
- [308] Qingji Guan, Yaping Huang, Zhun Zhong, Zhedong Zheng, Liang Zheng, and Yi Yang. Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification. *arXiv preprint arXiv:1801.09927*, 2018.
- [309] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
- [310] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

- [311] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [312] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.
- [313] Xueying Chen, Rong Zhang, and Pingkun Yan. Feature fusion encoder decoder network for automatic liver lesion segmentation. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 430–433. IEEE, 2019.
- [314] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. CBAM: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.
- [315] Changlu Guo, Márton Szemenyei, Yugen Yi, Wenle Wang, Buer Chen, and Changqi Fan. SA-UNet: Spatial attention u-net for retinal vessel segmentation. In *2020 25th international conference on pattern recognition (ICPR)*, pages 1236–1242. IEEE, 2021.
- [316] Zhengxuan Zhao, Kaixu Chen, and Satoshi Yamane. CBAM-UNet++: easier to find the target with the attention module CBAM. In *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, pages 655–657. IEEE, 2021.
- [317] Caiyong Wang, Yong He, Yunfan Liu, Zhaofeng He, Ran He, and Zhenan Sun. Sclerasetnet: an improved u-net model with attention for accurate sclera segmentation. In *2019 International Conference on Biometrics (ICB)*, pages 1–8. IEEE, 2019.
- [318] Sun Li, Liu Zhao, Jiyun Li, and Qian Chen. Segmentation of hippocampus based on 3DUNet-CBAM model. In *2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEM-CSE)*, pages 595–599. IEEE, 2021.
- [319] Yuan Zhong. Polyp segmentation using fully convolutional neural network with dropout and CBAM. In *International Conference on Computing and Data Science*, pages 171–181. Springer, 2021.
- [320] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7794–7803, 2018.
- [321] Zhengyang Wang, Na Zou, Dinggang Shen, and Shuiwang Ji. Non-local U-Nets for biomedical image segmentation. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 6315–6322, 2020.
- [322] Ran Gu, Guotai Wang, Tao Song, Rui Huang, Michael Aertsen, Jan Deprest, Sébastien Ourselin, Tom Vercauteren, and Shaoting Zhang. Ca-net: Comprehensive attention convolutional neural networks for explainable medical image segmentation. *IEEE transactions on medical imaging*, 40(2):699–711, 2020.

-
- [323] Joaquin Quinonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence. *Dataset shift in machine learning*. Mit Press, 2022.
- [324] Carlo Pierpaoli. Quantitative brain MRI, 2010.
- [325] Cynthia M Stonnington, Geoffrey Tan, Stefan Klöppel, Carlton Chu, Bogdan Draganski, Clifford R Jack Jr, Kewei Chen, John Ashburner, and Richard SJ Frackowiak. Interpreting scan data acquired from multiple scanners: a study with alzheimer’s disease. *Neuroimage*, 39(3):1180–1185, 2008.
- [326] Hans-Jürgen Huppertz, Judith Kröll-Seger, Stefan Klöppel, Reinhard E Ganz, and Jan Kassubek. Intra-and interscanner variability of automated voxel-based volumetry based on a 3d probabilistic atlas of human cerebral structures. *Neuroimage*, 49(3):2216–2224, 2010.
- [327] Ehab A AlBadawy, Ashirbani Saha, and Maciej A Mazurowski. Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. *Medical physics*, 45(3):1150–1158, 2018.
- [328] Veronika Cheplygina, Marleen de Bruijne, and Josien PW Pluim. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical image analysis*, 54:280–296, 2019.
- [329] Mohsen Ghafoorian, Alireza Mehrtash, Tina Kapur, Nico Karssemeijer, Elena Marchiori, Mehran Pesteie, Charles RG Guttmann, Frank-Erik de Leeuw, Clare M Tempany, Bram van Ginneken, et al. Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 516–524. Springer, 2017.
- [330] Konstantinos Kamnitsas, Christian Baumgartner, Christian Ledig, Virginia Newcombe, Joanna Simpson, Andrew Kane, David Menon, Aditya Nori, Antonio Criminisi, Daniel Rueckert, et al. Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. In *International conference on information processing in medical imaging*, pages 597–609. Springer, 2017.
- [331] Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2021.
- [332] Yanyang Gu, Zongyuan Ge, C Paul Bonnington, and Jun Zhou. Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification. *IEEE journal of biomedical and health informatics*, 24(5):1379–1393, 2019.
- [333] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

- [334] Ehsan Hosseini-Asl, Robert Keynton, and Ayman El-Baz. Alzheimer’s disease diagnostics by adaptation of 3d convolutional network. In *2016 IEEE international conference on image processing (ICIP)*, pages 126–130. IEEE, 2016.
- [335] Sergi Valverde, Mostafa Salem, Mariano Cabezas, Deborah Pareto, Joan C Vilanova, Lluís Ramió-Torrentà, Àlex Rovira, Joaquim Salvi, Arnau Oliver, and Xavier Lladó. One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks. *NeuroImage: Clinical*, 21:101638, 2019.
- [336] Aaron Carass, Snehashis Roy, Amod Jog, Jennifer L Cuzzocreo, Elizabeth Magrath, Adrian Gherman, Julia Button, James Nguyen, Ferran Prados, Carole H Sudre, et al. Longitudinal multiple sclerosis lesion segmentation: resource and challenge. *NeuroImage*, 148:77–102, 2017.
- [337] Joris Roels, Julian Hennies, Yvan Saeys, Wilfried Philips, and Anna Kreshuk. Domain adaptive segmentation in volume electron microscopy imaging. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1519–1522. IEEE, 2019.
- [338] Xiaofeng Liu, Fangxu Xing, Nadya Shusharina, Ruth Lim, C-C Jay Kuo, Georges El Fakhri, and Jonghye Woo. ACT: Semi-supervised domain-adaptive medical image segmentation with asymmetric co-training. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 66–76. Springer, 2022.
- [339] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.
- [340] Qi Dou, Cheng Ouyang, Cheng Chen, Hao Chen, Ben Glocker, Xiahai Zhuang, and Pheng-Ann Heng. Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. *IEEE Access*, 7:99065–99076, 2019.
- [341] Xiahai Zhuang and Juan Shen. Multi-scale patch and multi-modality atlases for whole heart segmentation of MRI. *Medical image analysis*, 31:77–87, 2016.
- [342] Wenjun Yan, Yuanyuan Wang, Menghua Xia, and Qian Tao. Edge-guided output adaptor: Highly efficient adaptation module for cross-vendor medical image segmentation. *IEEE Signal Processing Letters*, 26(11):1593–1597, 2019.
- [343] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, 8(6):679–698, 1986.
- [344] Mathilde Bateson, José Dolz, Hoel Kervadec, Hervé Lombaert, and I Ben Ayed. Constrained domain adaptation for image segmentation. *IEEE Transactions on Medical Imaging*, 40(7):1875–1887, 2021.

- [345] Amir Gholami, Shashank Subramanian, Varun Shenoy, Naveen Himthani, Xiangyu Yue, Sicheng Zhao, Peter Jin, George Biros, and Kurt Keutzer. A novel domain adaptation framework for medical image segmentation. In *International MICCAI Brainlesion Workshop*, pages 289–298. Springer, 2018.
- [346] Tianyang Zhang, Jun Cheng, Huazhu Fu, Zaiwang Gu, Yuting Xiao, Kang Zhou, Shenghua Gao, Rui Zheng, and Jiang Liu. Noise adaptation generative adversarial network for medical image analysis. *IEEE transactions on medical imaging*, 39(4):1149–1159, 2019.
- [347] Bo Li, Xinge You, Jing Wang, Qinmu Peng, Shi Yin, Ruinan Qi, Qianqian Ren, and Ziming Hong. IAS-NET: Joint intraclassly adaptive GAN and segmentation network for unsupervised cross-domain in neonatal brain MRI segmentation. *Medical Physics*, 48(11):6962–6975, 2021.
- [348] Jian Chen, Yue Sun, Zhenghan Fang, Weili Lin, Gang Li, Li Wang, UNC UMN Baby Connectome Project Consortium, et al. Harmonized neonatal brain MR image segmentation model for cross-site datasets. *Biomedical Signal Processing and Control*, 69:102810, 2021.
- [349] Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng Ann Heng. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. *IEEE transactions on medical imaging*, 39(7):2494–2505, 2020.
- [350] A Emre Kavur, N Sinem Gezer, Mustafa Barış, Sinem Aslan, Pierre-Henri Conze, Vladimir Groza, Duc Duy Pham, Soumick Chatterjee, Philipp Ernst, Savaş Özkan, et al. CHAOS challenge-combined (CT-MR) healthy abdominal organ segmentation. *Medical Image Analysis*, 69:101950, 2021.
- [351] Wenjun Yan, Yuanyuan Wang, Shengjia Gu, Lu Huang, Fuhua Yan, Liming Xia, and Qian Tao. The domain shift problem of medical image segmentation and vendor-adaptation by UNet-GAN. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 623–631. Springer, 2019.
- [352] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in bioinformatics*, 2018.
- [353] Mauricio Orbes-Arteaga, Thomas Varsavsky, Carole H. Sudre, Zach Eaton-Rosen, Lewis J. Haddow, Lauge Sørensen, Mads Nielsen, Akshay Pai, Sébastien Ourselin, Marc Modat, Parashkev Nachev, and M. Jorge Cardoso. Multi-domain adaptation in brain MRI through paired consistency and adversarial learning. In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*, pages 54–62, Cham, 2019. Springer International Publishing.
- [354] Kaiser Kushibar, Sergi Valverde, Sandra González-Villà, Jose Bernal, Mariano Cabezas, Arnau Oliver, and Xavier Lladó. Supervised domain adaptation for automatic sub-cortical brain structure segmentation with minimal user interaction. *Scientific reports*, 9(1):1–15, 2019.

- [355] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015.
- [356] Eric Kerfoot, Esther Puyol-Antón, Bram Ruijsink, Rina Ariga, Ernesto Zacur, Pablo Lamata, and Julia Schnabel. Synthesising images and labels between MR sequence types with CycleGAN. In *Domain Adaptation and Representation Transfer and Medical Image Learning with Less Labels and Imperfect Data*. Springer, 2019.
- [357] Andreas Schuh, Antonios Makropoulos, Emma C. Robinson, Lucilio Cordero-Grande, Emer Hughes, Jana Hutter, Anthony N. Price, Maria Murgasova, Rui Pedro A. G. Teixeira, Nora Tusor, Johannes K. Steinweg, Suresh Victor, Mary A. Rutherford, Joseph V. Hajnal, A. David Edwards, and Daniel Rueckert. Unbiased construction of a temporally consistent morphological atlas of neonatal brain development. *bioRxiv*, 2018.
- [358] Irina Grigorescu, Lucilio Cordero-Grande, Dafnis Batalle, A. David Edwards, Joseph V. Hajnal, Marc Modat, and Maria Deprez. Harmonised segmentation of neonatal brain MRI: A domain adaptation approach. In *Medical Ultrasound, and Preterm, Perinatal and Paediatric Image Analysis*, pages 253–263, Cham, 2020. Springer International Publishing.
- [359] M. B. M. Ranzini, I. Groothuis, K. Kläser, M. J. Cardoso, J. Henckel, S. Ourselin, A. Hart, and M. Modat. Combining multimodal information for metal artefact reduction: An unsupervised deep learning framework. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 600–604, 2020.
- [360] Haofu Liao, Wei-An Lin, S Kevin Zhou, and Jiebo Luo. Adn: Artifact disentanglement network for unsupervised metal artifact reduction. *IEEE transactions on medical imaging*, 39(3):634–643, 2019.
- [361] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017.
- [362] Nicholas J Tustison, Brian B Avants, Philip A Cook, Gang Song, Sandhitsu Das, Niels van Strien, James R Stone, and James C Gee. The ANTs cortical thickness processing pipeline. In *Medical Imaging 2013: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 2013.
- [363] Ziv Yaniv, Bradley C. Lowekamp, Hans J. Johnson, and Richard Beare. SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research. *Journal of Digital Imaging*, 31(3):290–303, 2018.
- [364] Bradley Lowekamp, David Chen, Luis Ibáñez, and Daniel Blezek. The design of SimpleITK. *Frontiers in Neuroinformatics*, 7:45, 2013.
- [365] Nancy Bayley. *Bayley scales of infant and toddler development*. PsychCorp, Pearson, 2006.

- [366] Veit Sandfort, Ke Yan, Perry J Pickhardt, and Ronald M Summers. Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. *Scientific reports*, 9(1):1–9, 2019.
- [367] Zoltan Nagy, Hugo Lagercrantz, and Chloe Hutton. Effects of preterm birth on cortical thickness measured in adolescence. *Cerebral Cortex*, 21(2):300–306, 2011.
- [368] Raymond Pomponio, Guray Erus, Mohamad Habes, Jimit Doshi, Dhivya Srinivasan, Elizabeth Mamourian, Vishnu Bashyam, Ilya M Nasrallah, Theodore D Satterthwaite, Yong Fan, et al. Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *NeuroImage*, 208:116450, 2020.
- [369] Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation with multiple sources. *Advances in neural information processing systems*, 21:1041–1048, 2008.
- [370] Ruijia Xu, Ziliang Chen, Wangmeng Zuo, Junjie Yan, and Liang Lin. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3964–3973, 2018.
- [371] Xiaofeng Liu, Fangxu Xing, Georges El Fakhri, and Jonghye Woo. Self-semantic contour adaptation for cross modality brain tumor segmentation. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE, 2022.
- [372] Yabin Zhang, Haojian Zhang, Bin Deng, Shuai Li, Kui Jia, and Lei Zhang. Semi-supervised models are strong unsupervised domain adaptation learners. *arXiv preprint arXiv:2106.00417*, 2021.
- [373] Andreas Schuh. Computational models of the morphology of the developing neonatal human brain. 2017.
- [374] Irina Grigorescu, Alena Uus, Daan Christiaens, Lucilio Cordero-Grande, Jana Hutter, Dafnis Batalle, A David Edwards, Joseph V Hajnal, Marc Modat, and Maria Deprez. Uncertainty-aware deep learning based deformable registration. In *Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Perinatal Imaging, Placental and Preterm Image Analysis*, pages 54–63. Springer, 2021.
- [375] J-Donald Tournier, Robert Smith, David Raffelt, Rami Tabbara, Thijs Dhollander, Maximilian Pietsch, Daan Christiaens, Ben Jeurissen, Chun-Hung Yeh, and Alan Connelly. MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage*, 202:116137, 2019.
- [376] Diederik P. Kingma, Danilo J. Rezende, Shakir Mohamed, and Max Welling. Semi-supervised learning with deep generative models, 2014.
- [377] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.

- [378] Leslie N. Smith. Cyclical learning rates for training neural networks, 2015.
- [379] Jessica Dubois, Marianne Alison, Serena J Counsell, Lucie Hertz-Pannier, Petra S Hüppi, and Manon JNL Benders. MRI of the neonatal brain: a review of methodological challenges and neuroscientific advances. *Journal of Magnetic Resonance Imaging*, 53(5):1318–1343, 2021.
- [380] Laura Gui, Serafeim Loukas, François Lazeyras, PS Hüppi, Djalel Eddine Meskaldji, and C Borradori Tolsa. Longitudinal study of neonatal brain tissue volumes in preterm infants and their ability to predict neurodevelopmental outcome. *Neuroimage*, 185:728–741, 2019.
- [381] Oliver Gale-Grant, Sunniva Fenn-Moltu, Lucas GS França, Ralica Dimitrova, Daan Christiaens, Lucilio Cordero-Grande, Andrew Chew, Shona Falconer, Nicholas Harper, Anthony N Price, et al. Effects of gestational age at birth on perinatal structural brain development in healthy term-born babies. *Human Brain Mapping*, 43(5):1577–1589, 2022.
- [382] Jelle Veraart, Jan Sijbers, Stefan Sunaert, Alexander Leemans, and Ben Jeurissen. Weighted linear least squares estimation of diffusion MRI parameters: strengths, limitations, and pitfalls. *NeuroImage*, 81:335–346, 2013.
- [383] Nagesh Adluru, Hui Zhang, Andrew S Fox, Steven E Shelton, Chad M Ennis, Anne M Bartosic, Jonathan A Oler, Do PM Tromp, Elizabeth Zakszewski, James C Gee, et al. A diffusion tensor brain template for rhesus macaques. *Neuroimage*, 59(1):306–318, 2012.
- [384] Peter J Basser and Sinisa Pajevic. Statistical artifacts in diffusion tensor MRI (DT-MRI) caused by background noise. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 44(1):41–50, 2000.
- [385] Denis Le Bihan, Cyril Poupon, Alexis Amadon, and Franck Lethimonnier. Artifacts and pitfalls in diffusion MRI. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 24(3):478–488, 2006.
- [386] Hui Zhang, Paul A Yushkevich, and James C Gee. Deformable registration of diffusion tensor mr images with explicit orientation optimization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 172–179. Springer, 2005.
- [387] Jonathan O’Muirheartaigh, Emma C Robinson, Maximillian Pietsch, Thomas Wolfers, Paul Aljabar, Lucilio Cordero Grande, Rui PAG Teixeira, Jelena Bozek, Andreas Schuh, Antonios Makropoulos, et al. Modelling brain development to detect white matter injury in term and preterm born neonates. *Brain*, 143(2):467–479, 2020.