

Volumetric MRI Reconstruction from 2D Slices in the Presence of Motion

A dissertation in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy

written at the
Wellcome/EPSRC Centre for Interventional and Surgical Sciences
Centre for Medical Image Computing
Department of Medical Physics and Biomedical Engineering
University College London

supervised by
Prof. Sébastien Ourselin,
Prof. Tom Vercauteren,
Dr. David Atkinson,
and
Prof. Anna David

by
Michael Ebner

London, 15th July, 2019

©2019
Michael Ebner
All Rights Reserved

I, Michael Ebner, 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 work.

Abstract

Despite recent advances in acquisition techniques and reconstruction algorithms, magnetic resonance imaging (MRI) remains challenging in the presence of motion. To mitigate this, ultra-fast two-dimensional (2D) MRI sequences are often used in clinical practice to acquire thick, low-resolution (LR) 2D slices to reduce in-plane motion. The resulting stacks of thick 2D slices typically provide high-quality visualizations when viewed in the in-plane direction. However, the low spatial resolution in the through-plane direction in combination with motion commonly occurring between individual slice acquisitions gives rise to stacks with overall limited geometric integrity. In further consequence, an accurate and reliable diagnosis may be compromised when using such motion-corrupted, thick-slice MRI data.

This thesis presents methods to volumetrically reconstruct geometrically consistent, high-resolution (HR) three-dimensional (3D) images from motion-corrupted, possibly sparse, low-resolution 2D MR slices. It focuses on volumetric reconstructions techniques using inverse problem formulations applicable to a broad field of clinical applications in which associated motion patterns are inherently different, but the use of thick-slice MR data is current clinical practice. In particular, volumetric reconstruction frameworks are developed based on slice-to-volume registration with inter-slice transformation regularization and robust, complete-outlier rejection for the reconstruction step that can either avoid or efficiently deal with potential slice-misregistrations. Additionally, this thesis describes efficient Forward-Backward Splitting schemes for image registration for any combination of differentiable (not necessarily convex) similarity measure and convex (not necessarily smooth) regularization with a tractable proximal operator.

Experiments are performed on fetal and upper abdominal MRI, and on historical,

printed brain MR films associated with a uniquely long-term study dating back to the 1980s. The results demonstrate the broad applicability of the presented frameworks to achieve robust reconstructions with the potential to improve disease diagnosis and patient management in clinical practice.

Impact Statement

The algorithms and software for volumetric MRI reconstruction from 2D slices developed in this thesis have the potential to lead to benefits both inside and outside of academia in several ways.

Within the academic environment, the presented reconstruction framework for fetal brain MRI enables the investigation of fetal and neonatal brain development in high resolution. It sets the stage for developing more robust and accurate segmentation, parcellation, and image analysis tools that benefit from the high-resolution representation, eventually leading to more precise, quantitative measurements of cortical brain development. Furthermore, the ability to accurately reconstruct a geometrically consistent, volumetric representation from printed brain MR films, as demonstrated with a uniquely long-term brain MR study dating back to the 1980s, makes patient film data accessible to modern image processing techniques. In the context of longitudinal studies in which historical data is preserved only on film, this could be vital in the development and measurement of biomarkers to help understand disease progression of brain conditions such as multiple sclerosis. Moreover, further research on the presented numerical framework of Forward-Backward Splitting may lead to the development of physiologically more plausible registration frameworks because biologically meaningful tissue properties, such as elasticity and incompressibility, can be incorporated more efficiently.

Outside of academia, the work presented in this thesis has the potential to benefit clinical practice and healthcare. The developed fully-automated reconstruction framework for fetal brain MRI may unlock the clinical use of high-resolution visualizations in the standard anatomical planes for both normal and pathological cases. The high-resolution reconstructions obtained by this framework may aid more ac-

curate brain tissue measurement, which would help identify pathological changes associated with conditions such as spina bifida and other prognostically important brain changes. In the case of fetal surgery, the higher resolution could facilitate presurgical planning, leading to safer procedures with better surgical outcomes. In the context of upper abdominal studies, in particular in conjunction with magnetic resonance cholangio-pancreatography, the higher anatomical clarity of peri-ductal and extra-ductal upper abdominal soft tissue anatomy, which can be achieved using the proposed reconstruction framework, may lead to improved clinical diagnosis and aid in the management of cancer patients.

Finally, the public availability of the developed frameworks as open-source software packages^{1,2,3,4,5} allows straightforward and widespread access to the presented concepts. Indeed, researchers and clinical scientists from several hospitals, including University College London Hospitals (UK), Sheffield Teaching Hospitals (UK), University Hospitals Leuven (Belgium), and Bambino Gesù Children’s Hospital of Rome (Italy), and numerous research institutions, including University College London (UK), Imperial College London (UK), King’s College London (UK), Medical University of Vienna (Austria), École Polytechnique Fédérale de Lausanne (Switzerland), Technical University of Valencia (Spain), Indiana University (USA), and Zhejiang University (China), are currently using the developed algorithms. In particular, the availability of these software packages has led to the translation of the presented reconstruction frameworks into clinical research tools used in hospitals and has encouraged knowledge exchange between academic institutions. Furthermore, this could foster future collaboration with industry partners, such as medical imaging companies, which could lead to accelerated technology transfer and clinical translation.

¹<https://github.com/gift-surg/NiftyMIC>

²https://github.com/gift-surg/ITK_NiftyMIC

³<https://github.com/gift-surg/VolumetricReconstructionFromPrintedFilms>

⁴<https://github.com/gift-surg/NSoL>

⁵<https://github.com/gift-surg/SimpleReg>

Acknowledgements

I would like to express my deepest gratitude to a range of people who have made it possible to complete this thesis.

First, I would like to thank my supervisors Sébastien Ourselin, Tom Vercauteren, David Atkinson and Anna David for all their guidance and support over the last few years. In particular, I would like to thank Sébastien Ourselin for the environment and the opportunities he provided. You allowed me to venture into exciting new territory both within and outside of academia, which has been invaluable in shaping my understanding of this field. I am also deeply grateful to Tom Vercauteren, who has provided me with guidance and tremendous support at every point of this journey. I would not have been able to finish all this exciting work without your supervision.

Moreover, I am thankful to Simon Arridge, Felix Lucka, David Atkinson, Manil Chouhan, Marc Modat, Tom Doel, Dzhoshkun Shakir and Luis García-Peraza Herrera for sharing their thoughts and expertise with me. This thesis is built around the knowledge and insights you provided.

Furthermore, I would like to thank Premal Patel. Together with Jan Deprest, Anna David, Manil Chouhan and Michael Aertsen, your continuous support has been vital in developing these algorithms that are now used in a clinical context.

With regards to the finishing touches of this thesis, I would like to thank Erinna Bowman and Premal. I feel many parts of this thesis make a lot more sense after it passed your corrections.

I am deeply thankful for the fantastic team and individuals I have met both north and south the river. Because of Alejandro, Loïc, Premal, Luis, Daniele, Wenqi, Efthymios, Oeslle, Ioana, Sebastiano, Lucas, and so many more, my experience in London has been a tremendous joy with countless dear memories.

Finally, I would like to thank my family. I am eternally grateful for your unconditional support and love that has made all this possible for me.

List of Publications

Original Contributions in International Peer-reviewed Journals

1. **Ebner, M.**, Patel, P. A., Atkinson, D., Caselton, L., Firmin, L., Amin, Z., Bainbridge, A., De Coppi, P., Taylor, S. A., Ourselin, S., Chouhan, M. D., and Vercauteren, T. (2019a). Super-resolution for upper abdominal MRI: Acquisition and post-processing protocol optimization using brain MRI control data and expert reader validation. *Magnetic Resonance in Medicine*, page mrm.27852.
2. **Ebner, M.**, Chung, K. K., Prados, F., Cardoso, M. J., Chard, D. T., Vercauteren, T., and Ourselin, S. (2018a). Volumetric reconstruction from printed films: Enabling 30 year longitudinal analysis in MR neuroimaging. *NeuroImage*, 165:238–250.
3. Xie, Y., Thom, M., **Ebner, M.**, Wykes, V., Desjardins, A., Miserocchi, A., Ourselin, S., McEvoy, A. W., and Vercauteren, T. (2017). Wide-field spectrally resolved quantitative fluorescence imaging system: toward neurosurgical guidance in glioma resection. *Journal of Biomedical Optics*, 22(11):1–14.

Submissions Under Review

1. **Ebner, M.**, Wang, G., Li, W., Aertsen, M., Patel, P. A., Aughwane, R., Melbourne, A., Doel, T., Dymarkowski, S., De Coppi, P., David, A. L., Deprest, J., Ourselin, S., and Vercauteren, T. (2019c). An Automated Framework for Lo-

calization, Segmentation and Super-Resolution Reconstruction of Fetal Brain MRI. *Under Submission (NeuroImage)*.⁶

2. Chung, K. K., Altmann, D., Barkhof, F., Miszkiel, K., Fisniku, L., Brex, P. A., O’Riordan, J., **Ebner, M.**, Prados, F., Cardoso, M. J., Vercauteren, T., Ourselin, S., Thompson, A., Ciccarelli, O., and Chard, D. T. (2019). A thirty year clinical and MRI observational study of multiple sclerosis and clinically isolated syndromes. *Under Revision (Annals of Neurology)*.
3. Ranzini, M. B. M., Henckel, J., **Ebner, M.**, Cardoso, M. J., Vercauteren, T., Ourselin, S., Hart, A., and Modat, M. (2019). Hip arthroplasty failure assessment using multimodal imaging joint segmentation. *Under Revision (Computer Methods and Programs in Biomedicine)*.

Full-Length Articles in Selective International Peer-reviewed Conferences with Published Proceedings

1. Fidon, L., **Ebner, M.**, Garcia-Peraza-Herrera, L. C., Modat, M., Ourselin, S., and Vercauteren, T. (2019). Incompressible image registration using divergence-conforming B-splines. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*.
2. **Ebner, M.**, Wang, G., Li, W., Aertsen, M., Patel, P. A., Aughwane, R., Melbourne, A., Doel, T., David, A. L., Deprest, J., Ourselin, S., and Vercauteren, T. (2018c). An Automated Localization, Segmentation and Reconstruction Framework for Fetal Brain MRI. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2018*, pages 313–320. Springer.⁷
3. **Ebner, M.**, Modat, M., Ferraris, S., Ourselin, S., and Vercauteren, T. (2018b). Forward-backward splitting in deformable image registration: A demons approach. In *2018 IEEE 15th International Symposium on Biomedical Imaging*

⁶Michael Ebner and Guotai Wang contributed equally.

⁷Michael Ebner and Guotai Wang contributed equally.

(*ISBI 2018*), pages 1065–1069, Washington, D.C. IEEE.

Articles in International Peer-reviewed Workshops with Published Proceedings

1. Ranzini, M. B. M., **Ebner, M.**, Cardoso, M. J., Fotiadou, A., Vercauteren, T., Henckel, J., Hart, A., Ourselin, S., and Modat, M. (2018). Joint Multi-modal Segmentation of Clinical CT and MR from Hip Arthroplasty Patients. In *MICCAI Workshop on Computational Methods and Clinical Applications in Musculoskeletal Imaging (MSKI) 2017*, volume 10734 LNCS, pages 72–84. Springer.
2. **Ebner, M.**, Chouhan, M., Patel, P. A., Atkinson, D., Amin, Z., Read, S., Punwani, S., Taylor, S., Vercauteren, T., and Ourselin, S. (2017a). Point-Spread-Function-Aware Slice-to-Volume Registration: Application to Upper Abdominal MRI Super-Resolution. In Zuluaga, M. A., Bhatia, K., Kainz, B., Moghari, M. H., and Pace, D. F., editors, *Reconstruction, Segmentation, and Analysis of Medical Images. RAMBO 2016*, volume 10129 of *Lecture Notes in Computer Science*, pages 3–13. Springer International Publishing.

Abstracts in Conferences

1. **Ebner, M.**, Patel, P. A., Atkinson, D., Caselton, L., Taylor, S. A., Bainbridge, A., Ourselin, S., Chouhan, M., and Vercauteren, T. (2019b). Reconstruction-based Super-Resolution for High-Resolution Abdominal MRI: A Preliminary Study. In *International Society for Magnetic Resonance in Medicine (ISMRM)*.
2. Melbourne, A., Wang, G., **Ebner, M.**, Vercauteren, T., Rosenbaum, D., Kasirer, Y., Wasertil, N., Ben-David, E., and Bin-Nun, A. (2019). T1 and T2 Weighted Image Segmentation from 1.0T Neonatal MRI. In *International Society for Magnetic Resonance in Medicine (ISMRM)*.

-
3. Chouhan, M., **Ebner, M.**, Patel, P. A., Atkinson, D., Firmin, L., Amin, Z., De Coppi, P., Ourselin, S., Vercauteren, T., and Taylor, S. A. (2019). Expert-reader validation of optimised Super-Resolution Reconstruction for upper abdominal MRI. In *European Society of Gastrointestinal and Abdominal Radiology (ESGAR)*.
 4. **Ebner, M.**, Xie, Y., Wykes, V., Miserocchi, A., McEvoy, A., Ourselin, S., and Vercauteren, T. (2018d). Spatial Regularisation based Reconstruction of Quantitative Fluorescence Imaging. In *Biophotonics Congress: Biomedical Optics Congress 2018 (Microscopy/Translational/Brain/OTS)*, page JTU3A.8, Hollywood, Florida United States. OSA.⁸
 5. **Ebner, M.**, Parker, K., Vercauteren, T., Ourselin, S., Wassertheurer, S., Hughes, A., and Hametner, B. (2017c). Reservoir pressure separation at brachial, carotid and radial arteries: A quantitative comparison and evaluation. *Artery Research*, 20:54.
 6. **Ebner, M.**, Parker, K., Vercauteren, T., Ourselin, S., Wassertheurer, S., Hughes, A., and Hametner, B. (2017b). Calculating reservoir pressure with or without flow information: Similarity and algorithmic sensitivity at radial artery. *Artery Research*, 20:78–79.
 7. **Ebner, M.**, Aertsen, M., Pratt, R., David, A. L., Pandya, P. P., Deprest, J., Kendall, G., Klusmann, M., Humphries, P., Hewitt, R., Atkinson, D., Vercauteren, T., and Ourselin, S. (2016). Volumetric MRI Reconstruction of Trachea in Fetuses with Complex Fetal Neck Masses. In *Proceedings of the Institute of Physics and Engineering in Medicine (IPEM) Conference on Fetal, Neonatal and Paediatric MR Imaging: Techniques and Applications*, Leeds, United Kingdom.

⁸Michael Ebner and Xijing Xie contributed equally.

Publications associated with Conference Workshop

Organization

1. Cardoso, M. J., Arbel, T., Melbourne, A., Bogunovic, H., Moeskops, P., Chen, X., Schwartz, E., Garvin, M., Robinson, E., Trucco, E., **Ebner, M.**, Xu, Y., Makropoulos, A., Desjardin, A., and Vercauteren, T., editors (2017). *Fetal, Infant and Ophthalmic Medical Image Analysis*, volume 10554 of *Lecture Notes in Computer Science*. Springer International Publishing, Cham.

Table of Contents

Abstract	5
Impact Statement	7
Acknowledgements	9
List of Publications	11
List of Figures	21
List of Tables	27
1. Introduction	29
1.1. MRI in the Presence of Motion	31
1.1.1. The Effect of Motion on Resolution in MRI	31
1.1.2. Fetal Magnetic Resonance Imaging	33
1.1.3. Magnetic Resonance Cholangio-Pancreatography	37
1.1.4. Historical MRI Acquisitions on Printed Films	41
1.2. Super-Resolution Reconstruction: Addressing Clinical Needs of Higher Resolution in MRI	44
1.3. Thesis Contributions	48
1.3.1. Super-Resolution Reconstruction of Fetal Brain MRI	48
1.3.2. Super-Resolution Reconstruction of Upper Abdominal MRI	50
1.3.3. Volumetric Reconstruction from Printed Films	51
1.3.4. Forward-Backward Splitting in Deformable Image Registration	52
1.3.5. Publicly Available Open-Source Implementations	53

1.4. Thesis Organisation	53
2. Super-Resolution Reconstruction of Fetal Brain MRI	57
2.1. Introduction	59
2.2. An Automated Framework for Localization, Segmentation and Super-Resolution Reconstruction of Fetal Brain MRI	68
2.2.1. Localization based on Coarse Segmentation	69
2.2.2. Fine Segmentation	71
2.2.3. Robust HR Volume Reconstruction Framework	73
2.3. Experiments and Results	79
2.3.1. Data	79
2.3.2. Implementation Details	80
2.3.3. Localization Results	83
2.3.4. Segmentation Results	84
2.3.5. Outlier-Robust SRR Results	86
2.4. Discussion	95
3. Super-Resolution Reconstruction of Upper Abdominal MRI	105
3.1. Introduction	107
3.2. Motion Correction and Volumetric Reconstruction for Upper Abdominal MRI	111
3.2.1. SRR using Reference-Guided Multimodal Deformable Motion Correction	113
3.2.2. Outlier-robust SRR using Monomodal Rigid Motion Correction	115
3.2.3. Point-Spread-Function-Aware Slice-to-Volume Registration . .	116
3.3. Point-Spread-Function-Aware Slice-to-Volume Registration: Application to Upper Abdominal MRI Super-Resolution	117
3.3.1. Data, Evaluation Methodology and Results	118
3.3.2. Discussion	124

3.4. Super-Resolution for Upper Abdominal MRI: Acquisition and Post-Processing Protocol Optimisation	125
3.4.1. Image Acquisition	126
3.4.2. Data, Evaluation Methodology and Results	129
3.4.3. Discussion	146
4. Volumetric Reconstruction from Printed Films: Enabling 30 Year Longitudinal Analysis in MR Neuroimaging	153
4.1. Introduction	155
4.2. Volumetric Reconstruction from Printed MR Films	161
4.2.1. Semi-Automatic Slice Extraction	164
4.2.2. Joint Regularised Motion and Intensity Correction Model . .	165
4.2.3. Initialisation of Volumetric Reconstruction Algorithm	166
4.2.4. Global Intensity Correction	167
4.2.5. Refinement of Volumetric Reconstruction	167
4.3. Data, Evaluation Methodology and Results	169
4.3.1. Data	169
4.3.2. Parametrisation of Volumetric Reconstruction Pipeline	170
4.3.3. Evaluation Methodology	172
4.3.4. Results	174
4.4. Discussion	181
5. Forward-Backward Splitting in Deformable Image Registration: A Demons Approach	187
5.1. Introduction	189
5.2. Forward-Backward Splitting for Image Registration	191
5.2.1. Forward-Backward Splitting Methods	191
5.2.2. Forward Step in Image Registration	193
5.2.3. Proximal Step in Image Registration	193
5.3. Experiments	195
5.3.1. Circle to C	195

5.3.2. 3D Anatomical MRI	196
5.4. Discussion	197
6. Conclusions and Future Work	201
6.1. Conclusions	201
6.2. Discussion and Future Work	203
Appendices	207
A. Super-Resolution Reconstruction of Fetal Brain MRI	209
A.1. Evaluation of Volumetric Self-Consistency	209
A.2. Clinical Evaluation	209
B. Super-Resolution Reconstruction of Upper Abdominal MRI	213
B.1. Optimisation Control Studies for Brain MRI SRR	213
B.2. Optimisation Studies for Upper Abdominal MRI SRR	215
Bibliography	221

List of Figures

1	Introduction	29
1.1.	Visualization of typical fetal MRI data acquired at 1.5 T	33
1.2.	Visualization of typical MR data in MRCP studies showing the anatomy of the biliary tree acquired at 1.5 T	40
1.3.	Scanned MR Film of Multiple Sclerosis/Clinically Isolated Syndrome subject acquired at the National Hospital for Neurology and Neurosurgery in 1986	42
1.4.	Longitudinal proton density scans of one subject followed up since the 1980s.	43
2	Super-Resolution Reconstruction of Fetal Brain MRI	57
2.1.	Chiari II malformation with ventriculomegaly	59
2.2.	Three example stacks of fetal brain MRI	62
2.3.	The proposed fully automatic framework for fetal brain reconstruction	68
2.4.	The proposed fetal brain localization method	69
2.5.	The proposed fetal brain segmentation method	71
2.6.	Outlier-robust super-resolution reconstruction framework for fetal brain MRI	74
2.7.	Volumetric reconstructions in patient-specific orientation and standard anatomical planes	80
2.8.	Visual comparison of different methods for fetal brain localization . .	83
2.9.	Visual comparison of different methods for fetal brain segmentation .	85

2.10. Original SVR registration parameter values and the outcome after using GPR	87
2.11. Comparison of retained slices using original SVR or SVR combined with GPR	88
2.12. Impact of outlier threshold on slice retention and slice similarities . .	89
2.13. Impact of outlier threshold for different inputs masks	90
2.14. Qualitative comparison of reconstruction methods and obtained masks in the template space	91
2.15. Qualitative comparison of reconstruction methods in the subject space	92
2.16. Quantitative evaluation of slice similarities for different reconstruction methods	94
2.17. Summary of clinical evaluation	95
2.18. Qualitative comparison of reconstruction methods in the template space (post-surgical spina bifida)	96
2.19. Qualitative comparison of reconstruction methods in the template space (post-surgical spina bifida)	97
2.20. Qualitative comparison of reconstruction methods in the template space (normal subject)	98
2.21. Qualitative comparison of reconstruction methods in the template space (pre-surgical spina bifida)	99
2.22. Qualitative comparison of reconstruction methods in the template space (pre-surgical spina bifida)	100
3 Super-Resolution Reconstruction of Upper Abdominal MRI	105
3.1. T2W brain and abdominal data	114
3.2. Evaluation of the residuals for all subjects and modes of our proposed reconstruction framework.	120
3.3. Visualization of original HASTE data and the field of view for the SR reconstruction algorithms.	122

3.4. Qualitative comparison between linearly resampled original HASTE data and reconstructions obtained by BTK, IRTK and our proposed approach.	123
3.5. Visualization of acquisition planes of SST2W images	127
3.6. Images obtained by extended MRCP protocol for abdomen and brain anatomies	128
3.7. Ground-truth (HR T2W) similarities for static and reference-guided SRR outcomes for the quasi-static control brain experiment whereby each of the seven subjects is assigned a different marker	134
3.8. Qualitative comparison of the static and reference-guided SRR outcome of one subject for various input data scenarios in the sagittal view	135
3.9. Self-consistency evaluation given by projected similarities for all slices of obtained abdominal SRRs for an increasing number of input stacks for different motion-correction strategies summarised for all eight subjects	138
3.10. Qualitative comparison between the SRR approaches using either two or six input stacks	139
3.11. Clinical evaluation by four radiologists for the abdomen of the third experiment for all eight subjects	141
3.12. Evaluation of ground-truth (HR T2W) similarities for PSF-aware registration	142
3.13. Ground-truth (HR T2W) similarities for first-order Tikhonov and isotropic Total Variation regularization SRR outcomes for the quasi-static brain experiment involving seven subjects	147
3.14. Qualitative comparison between first-order Tikhonov and isotropic Total Variation regularization in the final reconstruction step	148
3.15. T2W brain and abdominal data	150

4	Volumetric Reconstruction from Printed Films: Enabling 30 Year Longitudinal Analysis in MR Neuroimaging	153
4.1.	Scanned MR Films of MS/Clinically Isolated Syndrome (CIS) subjects from 1985 to 1991	155
4.2.	Overview of volumetric reconstruction framework for historical MR film data	162
4.3.	Semi-Automatic Slice Extraction	164
4.4.	Initialisation of the volumetric reconstruction algorithm	166
4.5.	Refinement of the volumetric reconstruction algorithm	168
4.6.	Volumetric reconstruction framework validation pipeline	172
4.7.	Similarity measures evaluated at the ground-truth brain for each subject separately	174
4.8.	Similarity measures evaluated at the ground-truth brain for each individual subject	175
4.9.	Visual comparison between reconstruction results and original data .	178
4.10.	Qualitative comparison of three longitudinal scans with the baseline scan acquired in 1986	179
4.11.	Reconstructed sample to show the importance of inter-slice regularisation for motion correction to obtain high-quality volumetric reconstructions	182
5	Forward-Backward Splitting in Deformable Image Registration: A Demons Approach	187
5.1.	Convergence comparison of solvers for the non-linear registration of the classical “Circle to C” experiment	195
5.2.	3D-Comparison of FISTA and iPiano against the conjugate gradient solver based on overall 1180 registrations	196
6	Conclusions and Future Work	201

A	Super-Resolution Reconstruction of Fetal Brain MRI	209
A.1.	Slice similarities after the final SVR-SRR iteration	210
A.2.	Summary of clinical evaluation for anatomical clarity scores	211
A.3.	Summary of clinical evaluation for SRR quality scores	211
B	Super-Resolution Reconstruction of Upper Abdominal MRI	213
B.1.	Qualitative comparison of the static and reference-guided SRR outcome of one subject for various input data scenarios	214
B.2.	Ground-truth (HR T2W) similarities for static and reference-guided SRR outcomes for the quasi-static brain experiment involving seven subjects	215
B.3.	Ground-truth (HR T2W) similarities for the quasi-static brain experiment for all registration/motion correction strategies as an extension to Figure B.2	216
B.4.	Self-consistency evaluation given by projected similarities for all slices of obtained abdominal SRRs for an increasing number of input stacks for different motion-correction strategies summarised for all eight subjects	218

List of Tables

1	Introduction	29
2	Super-Resolution Reconstruction of Fetal Brain MRI	57
2.1.	Information of the datasets used for the experiments	78
2.2.	Quantitative evaluation of different methods for fetal brain localization	83
2.3.	Quantitative evaluation of fetal brain segmentation	84
2.4.	Assessment of the robustness of the proposed template-space alignment approach	93
3	Super-Resolution Reconstruction of Upper Abdominal MRI	105
3.1.	Summary of clinical evaluation averaged over all four subjects.	121
3.2.	Image acquisition protocol used in this volunteer study for both abdominal and control brain anatomies	129
3.3.	Ground-truth (HR T2W) NCC-similarities of obtained quasi-static control brain SRRs for an increasing number of input stacks for different motion correction strategies summarised for all seven subjects	135
3.4.	Clinical evaluation by two radiologists for abdomen averaged over all eight subjects for two separate experiments	137
3.5.	Evaluation of ground-truth (HR T2W) similarities for PSF-aware registration	143
3.6.	Evaluation of slice-similarities for PSF-aware registration	143

3.7. Ground-truth (HR T2W) similarities of obtained quasi-static control brain SRRs using first-order Tikhonov and isotropic Total Variation regularization SRR outcomes	146
3.9. Typical computational times to create a HR visualization of the biliary tree split into motion correction and volumetric reconstruction processing times	149
4 Volumetric Reconstruction from Printed Films: Enabling 30 Year Longitudinal Analysis in MR Neuroimaging	153
4.1. Summary of similarity measures evaluated at the ground-truth brain stated as mean and standard deviation for all 20 subjects	175
4.2. Summary of blinded, clinical evaluation averaged over all 20 subjects	176
4.3. Inter-slice regularisation: Summary of similarity measures evaluated at the ground-truth brain stated as mean and standard deviation for all 20 subjects	181
5 Forward-Backward Splitting in Deformable Image Registration: A Demons Approach	187
6 Conclusions and Future Work	201
A Super-Resolution Reconstruction of Fetal Brain MRI	209
B Super-Resolution Reconstruction of Upper Abdominal MRI	213
B.1. Ground-truth (HR T2W) similarities of obtained quasi-static control brain SRRs for an increasing number of input stacks for different motion correction strategies summarized for all seven subjects	217
B.3. Projected slice similarity evaluation of obtained abdominal SRRs for an increasing number of input stacks for different motion correction strategies summarized for all eight subjects	219

CHAPTER 1.

Introduction

Table of Contents

1.1. MRI in the Presence of Motion	31
1.1.1. The Effect of Motion on Resolution in MRI	31
1.1.2. Fetal Magnetic Resonance Imaging	33
1.1.3. Magnetic Resonance Cholangio-Pancreatography	37
1.1.4. Historical MRI Acquisitions on Printed Films	41
1.2. Super-Resolution Reconstruction: Addressing Clinical Needs of Higher Resolution in MRI	44
1.3. Thesis Contributions	48
1.3.1. Super-Resolution Reconstruction of Fetal Brain MRI	48
1.3.2. Super-Resolution Reconstruction of Upper Abdominal MRI	50
1.3.3. Volumetric Reconstruction from Printed Films	51
1.3.4. Forward-Backward Splitting in Deformable Image Regis- tration	52
1.3.5. Publicly Available Open-Source Implementations	53
1.4. Thesis Organisation	53

Since the early 1980s, when magnetic resonance imaging (MRI) first became clinically available, the advances in MR scanners and the development of bespoke magnetic resonance (MR) acquisition protocols have led to a widespread availability of MRI in clinical practice. MRI offers remarkable soft tissue contrast and the benefit

of non-ionizing radiation, making it a safe and highly valuable imaging modality for disease diagnosis and presurgical planning for a wide range of clinical applications. Nonetheless, the high sensitivity of MRI leads to specific challenges. On the one hand, it offers the possibility to obtain images with highly tunable soft tissue contrast indispensable for clinical diagnosis. On the other hand, it makes MR tremendously challenging in the presence of motion and generally results in lower-resolution and possibly motion-artefact-affected images. However, sufficient spatial resolution is key to visualizing small anatomical structures and is essential for a reliable clinical assessment.

The use of firmly established ultra-fast two-dimensional (2D) MRI in current clinical practice mitigates the problem of motion by acquiring a sequence of thick 2D slices spanning the entire anatomy of interest. The advantages of such ultra-fast 2D MRI are the ability to largely “freeze” the effect of motion as well as the provision of excellent soft tissue contrast with remarkably high in-plane resolution. However, a balance has to be struck between a short scanning time (to avoid motion artefacts) and the signal-to-noise ratio, which must be maintained at an acceptable level. With motion commonly occurring between each slice acquisition, the resulting “stack” of thick slices offers an anatomical visualization with limited geometric integrity, in which adjacent slices do not necessarily visualize adjacent anatomy. Moreover, the large slice thickness comes at a cost: small tissue structures with fine anatomical details cannot be captured. In further consequence, an accurate and reliable diagnosis based on such motion-corrupted, thick-slice MRI data may be compromised [Gholipour et al., 2014, Brix et al., 2016, Basaran et al., 2008, Griffin et al., 2012].

Recent advances in image post-processing have demonstrated the potential to increase the image resolution a posteriori by combining multiple MRI stacks of low-resolution 2D slices into a single, high-resolution 3D volume – in a method called *Super-Resolution Reconstruction*¹. Instead of acquiring a high-resolution 3D MR volume directly (which cannot be achieved due to motion and/or scanning time con-

¹Also referred to as Slice-to-Volume Reconstruction in this context, e.g. [Kainz et al., 2015b, Cordero-Grande et al., 2018].

siderations) *one can estimate* a high-resolution visualization from several, possibly motion-corrupted, low-resolution 2D slices. As a result, a more detailed anatomical visualization in higher resolution becomes possible which can lead to improved disease diagnosis and pre-surgical management. However, precise motion correction is vital to establish accurate inter-slice spatial correspondences.

The work in this thesis aims to provide robust methods to volumetrically reconstruct geometrically consistent, high-resolution 3D images from motion-corrupted, possibly sparse, low-resolution 2D MR slices. The thesis focuses on volumetric reconstruction techniques applicable to a broad field of clinical applications in which associated motion patterns are inherently different, but the use of thick-slice MR data is current clinical practice. In particular, a volumetric reconstruction framework is developed and applied to fetal (Chapter 2) and upper abdominal MRI (Chapter 3), and to historical, printed MR films associated with a uniquely long-term brain MR study dating back to the 1980s (Chapter 4). Each application is characterized by different motion patterns, including unpredictable, large motion (fetal MRI), breathing and bowel motion (abdominal MRI), and random motion (printed MR films).

1.1. MRI in the Presence of Motion

1.1.1. The Effect of Motion on Resolution in MRI

“A fundamental consideration in any MRI experiment is how to optimally balance image resolution, signal-to-noise ratio (SNR), and acquisition time. These three imaging parameters are highly interdependent: higher resolution allows one to observe smaller details, but typically reduces SNR, and/or increases imaging time. At the same time, a certain minimum level of SNR is required to distinguish the signal of interest from system noise, and scan time should be kept low, as MR imaging resources are limited, costly, and long scan times are uncomfortable for the patient, and induce motion artefacts in the images. Together, the three parameters span a space, and their theoretical relations are well

known.” [Plenge et al., 2012]

Indeed, the use of MR in the context of moving body organs has historically been difficult. Significant improvement of MRI hardware and software in recent years has led to the development of rapid imaging sequences to mitigate the challenging problem of motion. The type of underlying motion, however, may have a substantial impact on the preferred type of MR acquisition. Repetitive motion-patterns like cardiac or respiratory motion can be addressed by gated MR image acquisitions to reduce motion artefacts while maintaining relatively high resolution. A surrogate signal, like an echo-cardiogram or a respiratory signal, allows the synchronization of the image acquisition with the recurring relative position of the moving tissue. Therefore, the image acquisition is prolonged over several cycles until the entire image is constructed. MRI is particularly challenging in the presence of unpredictable or non-repetitive motion as is the case in, e.g., fetal MRI. There, typically, ultra-fast 2D MRI is performed in order to largely “freeze” motion for each slice acquisition. A short scanning time is critical to avoid motion artefacts and, together with SNR considerations, imposes a relatively low resolution on the imaging in practice. The resulting coarse spatial discretization, characterized by large, usually anisotropic, voxel sizes, is prone to obscuring small anatomical structures. Associated partial voluming effects are highly dependent on the directionality of the acquisition due to voxel anisotropy and severely impact an accurate representation of the underlying anatomy.

Overall, MR image quality and spatial resolution are affected primarily by the presence of motion. In consequence, the MR-based assessment in various clinical applications, including fetal and pediatric MRI, abdominal MRI and cardiac MRI, currently is constrained by the use of thick-slice data for clinical diagnosis. Information on small anatomical structures, relevant for disease diagnosis, may not be captured, limiting the diagnostic power of MRI in the presence of motion [Gholipour et al., 2014, Brix et al., 2016, Basaran et al., 2008, Griffin et al., 2012].

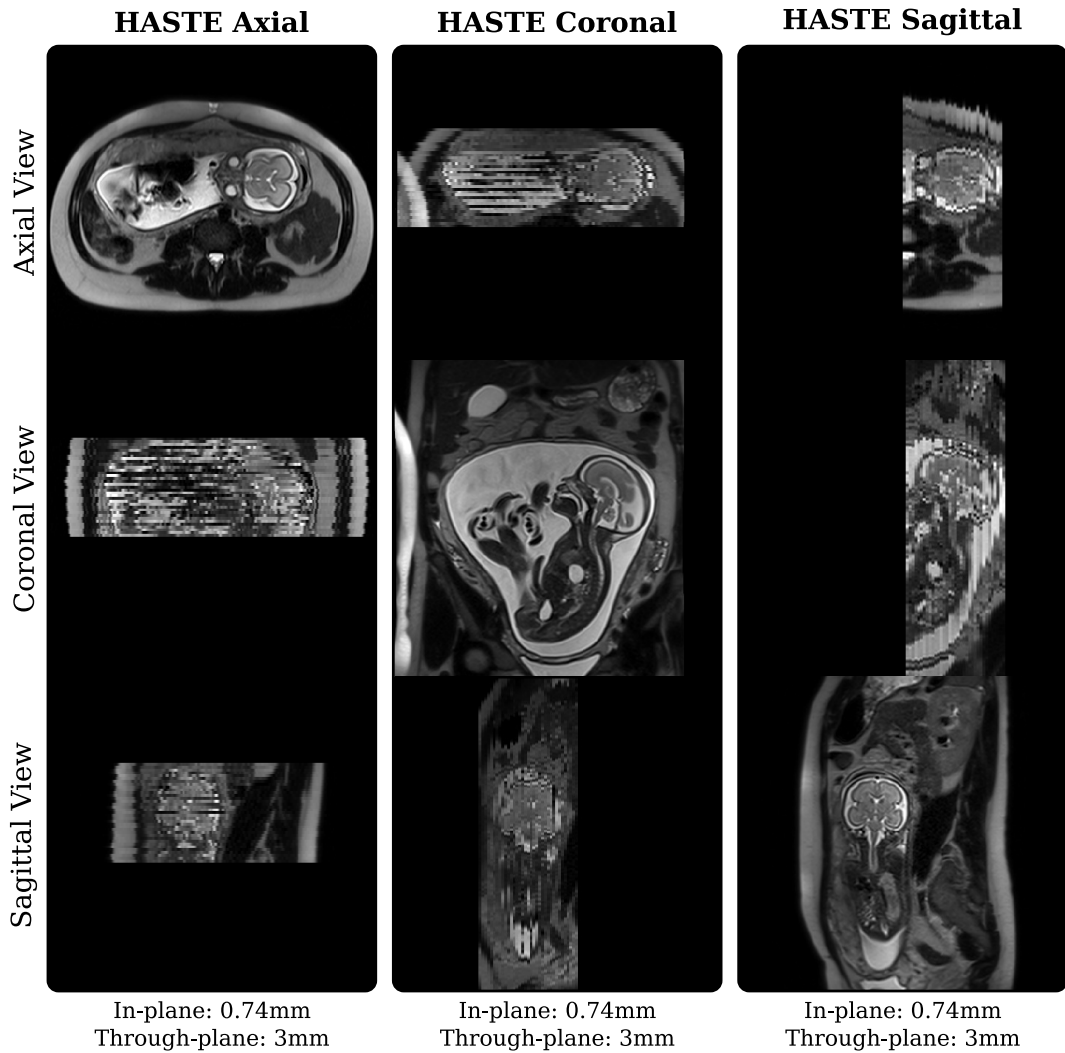


Figure 1.1.: Visualization of typical fetal MRI data acquired at 1.5 T (Data courtesy of Prof. Jan Deprest, KU Leuven). An interleaved scanning protocol was deployed where each stack was acquired using three packages of temporarily sequential but spatially separated slices. Maternal and fetal motion is visible throughout the HASTE stacks. Axial and coronal stacks show the effect of magnetization transfer where some slices have suffered from reduced signal intensity. The coronal stack is affected by wrapping artefacts.

1.1.2. Fetal Magnetic Resonance Imaging

Imaging technologies must meet strict safety standards in order to be utilized in the sensitive period of pregnancy. Thus, imaging modalities used in fetal examinations in clinical practice are typically constrained to ultrasound (US) and magnetic resonance

imaging (MRI), because these are considered safe due to their non-invasive and non-ionizing nature.

US remains the predominant modality employed during pregnancy to screen and diagnose fetal anomalies, mainly due to its cost-effectiveness, wide availability and portability, and real-time capabilities [Sepulveda et al., 2012]. However, US imaging is highly operator-dependent and the interpretation of certain US findings may be difficult even for skilled experts [Sepulveda et al., 2012, Benson and Bluth, 2008]. Moreover, US offers limited soft tissue contrast. Two-dimensional ultrasound (2D-US) in particular faces limitations in depicting fetal anatomy due its sensitivity to maternal habitus, abdominal scarring, the size and position of the fetus and its limitation at progressive ossification of the fetal skeleton which results in acoustic shadowing and oligohydramnios [Sepulveda et al., 2012, Benson and Bluth, 2008]. With the introduction of three-dimensional ultrasound (3D-US), volumetric acquisitions have been possible, giving rise to facilitated visualization of arbitrary sections and ultimately easier handling. Moreover, 3D-US allowed the storage of volumetric information that can be processed offline for further image analysis. However, 3D-US may only capture a limited field of view of the targeted organ, is considerably more expensive than 2D-US and remains highly operator-dependent [Sepulveda et al., 2012].

At present, fetal MRI is widely considered as a complementary tool to US according to [Garel, 2008]:

- (i) To reassure the parents of the normality of a fetal brain in the presence of normal US findings, but a positive family history of a cerebral abnormality that is recognised as being possibly overlooked by US, such as subtle abnormal gyration or posterior fossa anomalies.
- (ii) To clarify abnormalities detected by US and to add additional information that might have an influence on prognosis or management at birth.
- (iii) To add functional evaluation to the morphological US analysis.

The use of MRI has historically been difficult in fetal examinations due to the motion artefacts generated by maternal breathing and spontaneous fetal motion [Sepulveda et al., 2012]. Sedation of the fetus is highly controversial due to potential adverse effects on the developing brain and is generally prohibited [McPherson and Inder, 2017]. The development of rapid MR sequences has made it possible to mitigate this difficulty and has allowed MRI to gain momentum in clinical applications despite its relative cost and complexity [Prayer et al., 2004, Malamateniou et al., 2013, Saleem, 2014]. Fast imaging methods like gradient echo imaging, echo planar imaging, and fast spin-echo (FSE) imaging have been developed to collect images more rapidly and have substantially reduced imaging times compared to standard spin-echo imaging sequences [Dale et al., 2015]. In particular, T2-weighted ultra-fast FSE sequences allow sufficiently short scan times of less than one second, providing the ability to essentially “freeze” motion at each acquisition while providing excellent T2-contrast. The field of fetal MRI has since been dominated by the use of such single-shot T2-weighted (SST2W) sequences [Gholipour et al., 2014] with the associated vendor acronyms being

- HASTE for **H**Alf-fourier **S**ingle-shot **T**urbo spin **E**cho (Siemens)
- SSFSE for **S**ingle-**S**hot **F**ast **S**pin **E**cho (General Electric Medical Systems)
- SSHTSE for **S**ingle-**S**hot **H**alf-**F**ourier **T**urbo **S**pin **E**cho, UFSE for **U**ltra-**F**ast **S**pin **E**cho or TSE for **T**urbo **S**pin **E**cho (Philips)
- RARE for **R**apid **A**cquisition with **R**elaxation **E**nhancement (Bruker Instruments)

Currently, fetal MRI plays an important role in providing valuable information for disease diagnosis and ante-natal management and is especially useful for investigating suspected abnormalities detected during a US examination. Moreover, fetal MRI has been shown to yield improved diagnostic accuracy over US in a number of applications, such as the diagnosis of esophageal atresia [Ethun et al., 2014], lung volumetry [Story et al., 2017] and for various other conditions [Sepulveda et al., 2012].

Nevertheless, and despite attempts to further improve imaging speed with techniques like parallel imaging [Brown et al., 2014], MRI remains relatively slow and its application difficult in the context of motion. The need for short acquisition times to avoid motion artefacts in combination with the requirement to maintain reasonable signal-to-noise ratios leads to thick 2D slice acquisitions and ultimately to relatively low spatial resolution. In current clinical practice, typical SST2W images have $1 \times 1 \text{ mm}^2$ to $2 \times 2 \text{ mm}^2$ in-plane spatial resolution with slice thicknesses between 2 mm and 4 mm in fetal MRI [Gholipour et al., 2014]. Sequential slice acquisitions are performed with minimum delays to avoid inter-slice fetal motion, which results in an image represented as a stack of single 2D slices covering a selected volume of fetal anatomy. To keep the acquisition time and the impact of fetal motion low, the “cramming” of as many slices as possible into a specified period of time has been shown to be beneficial. However, this can lead to deleterious visual effects affecting the image quality due to amniotic fluid saturation, convective motion, and magnetization transfer between adjacent slices [Gholipour et al., 2014]. To avoid such slice crosstalk and spin history artefacts, slices are usually acquired in an interleaved manner, typically leading to two or three temporally sequential but spatially separated acquisitions. A typical total scan time of 20 s allows the acquisition of about 30 slices, which provides a coverage of about 12 cm of the fetus in the slice-select direction. Typically, several stacks are acquired orthogonal to each other in fetal MRI to obtain high in-plane resolution visualization in multiple planes in order to obtain sufficient anatomical coverage for improved diagnosis. However, with fetal and maternal motion commonly occurring during acquisition time, each resulting image stack provides a low-resolution visualization of the covered anatomy with overall limited geometric integrity where adjacent slices are not necessarily visualizing adjacent tissue anatomy (Figure 1.1). The high voxel anisotropy gives rise to substantial partial voluming effects and hides fine anatomical details valuable for reliable disease diagnosis. Moreover, optimal imaging planes for clinical assessment relative to the fetal anatomy, such as the fetal brain, cannot be guaranteed due to the unpredictable fetal motion. For longitudinal studies this may lead to difficulties

in assessing anatomical changes as it cannot be ensured that the structure of interest is visualised. Interactive, real-time imaging strategies have been proposed for more reliable slice plane selections to capture optimal imaging planes for increased diagnostic image quality [Brix et al., 2016]. However, while they represent promising approaches to obtain more accurate imaging planes for anatomical visualization, each associated slice is a low-resolution image and therefore suffers from substantial partial voluming.

1.1.3. Magnetic Resonance Cholangio-Pancreatography

The pancreatic ductal and biliary systems form part of the body’s digestive system to provide and conduct the pancreatic and bile juices (secretions) required for digestion. The production, storage, and secretion of the bile are guaranteed by a network of ducts interconnecting the liver, gallbladder, and pancreas. Several conditions and diseases can affect this digestive system. Gallstones can develop and lead to the obstruction of ducts and eventually give rise to severe inflammation at various points along the biliary tree, including inflammation of the gallbladder (cholelithiasis) and ducts of the biliary tree (choledocholithiasis) and pancreas (pancreatitis). In fact, gallstone disease is one of the most common and costly gastrointestinal (GI) tract disorders in the Western world with a prevalence of 10 % to 20 % [Li et al., 2017b, van Dijk et al., 2017]. In the United States, pancreatic cancer is the fourth leading cause of cancer-related deaths and is projected to become the second most deadly cancer in the near future [Yang et al., 2019, Siegel et al., 2019]. In the UK, it is the fifth most common cause of cancer death with less than 7% of affected people surviving more than five years [Scott and Jewell, 2019]. Imaging of the upper GI tract plays a vital role in early detection and treatment of such disorders. Common imaging modalities include endoscopic ultrasound (EUS), computed tomography (CT), endoscopic retrograde cholangio-pancreatography (ERCP) and magnetic resonance cholangio-pancreatography (MRCP), with all having their role in the diagnosis of biliary pathologies in current clinical practice [Walshe et al., 2016].

Endoscopic Ultrasound

EUS combines a high-frequency ultrasound probe with an endoscope and has been in use since the early 1980s in a variety of GI conditions. It is widely considered as safe with low rates of complications [Safari et al., 2016]. With a high diagnostic accuracy of about 95 %, EUS allows the reliable identification of bile duct stones (choledocholithiasis) [Chen et al., 2015, Makmun et al., 2017]. EUS also provides additional interventional capabilities, such as its use for guiding fine needle aspiration biopsy (EUS-FNA) for evaluating gastrointestinal and pulmonary malignancies [Williams et al., 1999]. However, EUS is highly operator-dependent, and its visualizations of the biliary tree may be incomplete or unsuccessful [Chen et al., 2015]. In particular, the assessment of the distal common bile duct may be limited where bowel gas, debris, fluid in the duodenum and obesity can degrade the image quality [Diwanji et al., 2016].

Computed Tomography

CT has the advantage of being non-invasive and allows imaging at high resolution with isotropic sub-millimeter voxels [Ringe and Wacker, 2015]. It is the standard imaging technique for detection of acute pancreatitis [Carroll et al., 2007]. However, the intravenous iodinated contrast agent required to differentiate pathology from normal anatomy on CT is nephrotoxic and can cause an allergic reaction. Moreover, it carries the risk of radiation exposure and fails at demonstrating two important pathologies of biliary stones and biliary strictures [Diwanji et al., 2016].

Endoscopic Retrograde Cholangio-Pancreatography

ERCP was first used in 1968 [McCune et al., 1968] and was quickly accepted as a technique for evaluating pancreatobiliary disease [Chathadi et al., 2015]. In fact, it has become the gold standard for diagnosis of choledocholithiasis [Chen et al., 2015, Makmun et al., 2017]. The procedure combines endoscopy and fluoroscopy; the endoscope is inserted through the mouth, into the stomach, and to the duodenum where contrast medium is injected into the biliary tree and pancreatic ducts for fur-

ther visualization via x-ray. As a therapeutic procedure, ERCP is particularly useful for removing biliary obstructions due to choledocholithiasis, or, in conjunction with bile duct stenting or biliary sphincterotomy, for treating bile leaks [Chathadi et al., 2015]. However, ERCP is highly dependent on the operator’s skill and experience and can cause serious complications because of its invasive nature. Moreover, it is associated with several complications such as pancreatitis, cholangitis, bleeding, and bowel perforation [Makmun et al., 2017], in addition to 1 % to 7 % morbidity and 0.2 % to 1 % mortality [Diwanji et al., 2016]. Due to its risks, ERCP has evolved from a diagnostic procedure to one that is almost exclusively therapeutic [Chathadi et al., 2015, Makmun et al., 2017]. Therefore, EUS and MRCP have become the modalities of choice for diagnosing choledocholithiasis [Makmun et al., 2017].

Magnetic Resonance Cholangio-Pancreatography

The use of MR for cholangiography was first described in [Wallner et al., 1991] and takes advantage of the inherent contrast-related properties of fluid in the biliary and pancreatic ducts [Barish et al., 1999]. The development of novel and faster MR imaging sequences in recent years has enabled the visualization of biliary and pancreatic systems with excellent image quality, which previously could only be provided by ERCP [Altun et al., 2016]. Moreover, MRCP provides images with inherently high soft tissue contrast, allows for a non-invasive examination without administration of anesthesia or intraductal or intravenous contrast agent, and is not operator dependent [Altun et al., 2016, Ringe and Wacker, 2015]. This makes MRCP a useful tool for use in almost all patients, including infants and those with allergies to iodine-based contrast materials [Safari et al., 2016]. As a diagnostic procedure, it is now increasingly replacing ERCP due to its non-invasiveness and excellent image quality combined with its comparable accuracy for a broad spectrum of benign and malignant pancreatic biliary ductal diseases [Barish et al., 1999, Chen et al., 2015, Ringe and Wacker, 2015]. Nevertheless, a comparison against ERCP as a diagnostic tool is still subject to ongoing investigations [Aydelotte et al., 2015, Kanaan and Antaki, 2016, Parthasarathy et al., 2016, Engelbrecht et al., 2016, Aydelotte, 2016, Shiani

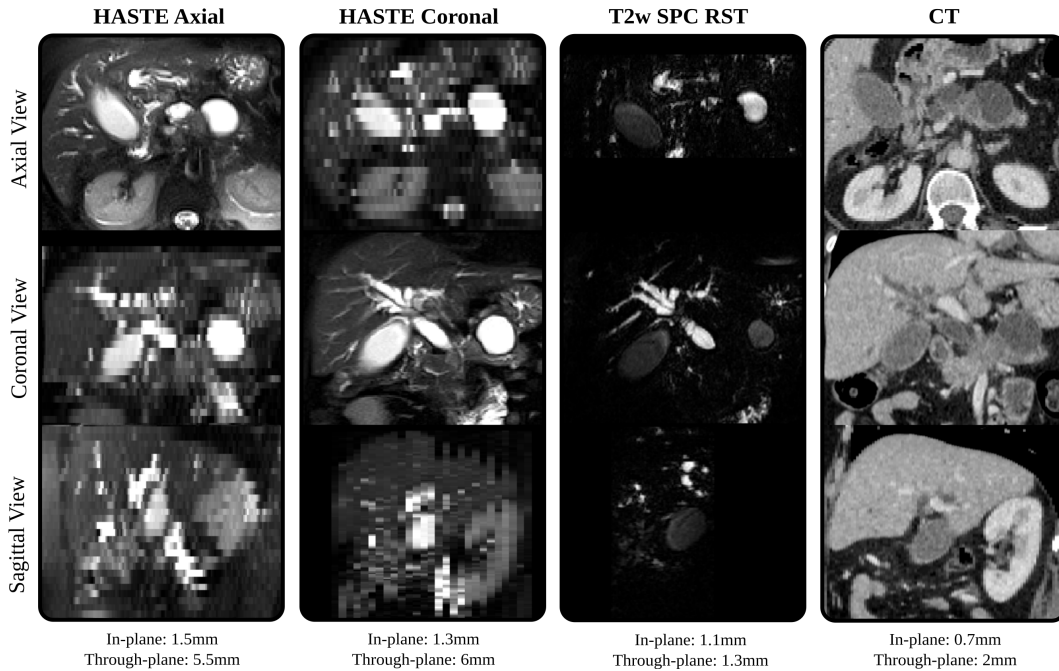


Figure 1.2.: Visualization of typical MR data in MRCP studies showing the anatomy of the biliary tree acquired at 1.5 T (Image adapted from [Ebner et al., 2017a]). Motion is visible throughout the HASTE stacks. The heavily T2-weighted volume (T2w SPC RST) has approximately five times higher resolution compared to the HASTE through-plane direction. However, the heavily T2-weighted volume loses valuable tissue contrast in the surrounding anatomy. Additionally acquired contrast-enhanced computed tomography (CT) image allows assessment in high resolution but does not have the inherent high soft tissue contrast of MRI and carries the risk of radiation exposure, iodinated contrast exposure and toxicity of kidneys.

et al., 2018, Akbar et al., 2018]. In contrast to ERCP, however, MRCP does not enable therapeutic maneuvers which remains a significant advantage of ERCP, which can perform therapeutic interventions at the time of initial diagnosis [Altun et al., 2016]. MRI of the upper GI anatomy remains challenging due to its susceptibility to motion and therefore the problem of complex motion during acquisition time, including respiratory and cardiac motion in addition to peristaltic bowel movement. Usually, a set of rapid T1- and T2-weighted sequences is acquired in current clinical protocols. The T1-weighted images are useful for the evaluation of duct walls and parenchymal lesions and can be acquired as T1-weighted gradient-echo sequences in a 2D or 3D fashion [Altun et al., 2016]. The T2-weighted images include a non-

breath-hold respiratory-triggered 3D turbo spin echo (TSE) and, typically, a coronal and an axial ultra-fast single-shot T2-weighted (SST2W) sequence at breath-hold (Figure 1.2). Respiratory-gating allows the acquisition of heavily T2-weighted images at high resolution of about 1 mm^3 isotropic voxel size. The signal from the pancreatico-biliary system appears hyper-intense, whereas the background tissue, such as hepatic and pancreatic tissue, peritoneal fat, fast-flowing blood, is either very low signal or signal void, resulting in excellent contrast and depiction of the pancreatico-biliary system [Altun et al., 2016]. Its downside, however, is a lack of surrounding structural tissue information, which is important for the assessment of peri-ductal and extra-ductal upper abdominal soft tissue pathology. The ultra-fast SST2W sequences require less than 1 s per slice acquisition and essentially “freeze” any physiological motion and, in contrast to the 3D TSE acquisition, capture information on peri- and extra-ductal anatomy. Acquired at breath-hold with durations of, typically, about 20 s a volumetric representation of the upper GI anatomy via a stack of (thick) 2D slices is possible. SNR-considerations prevent acquiring small slice thickness which, generally, leads to slices with slice thickness between 4 mm to 6 mm at an in-plane resolution of, approximately, 1 mm. Thus, the large voxel anisotropy manifests in considerable partial voluming effects and prevents the diagnosis of small vessels. In fact, and in contrast to ERCP, MRCP is known to miss gallstones smaller than 4 mm, small ampullary lesions and ductal strictures [Carroll et al., 2007, Safari et al., 2016].

1.1.4. Historical MRI Acquisitions on Printed Films

After the discovery of nuclear magnetic resonance by Purcell, Torrey and Pound in 1945 as described in the seminal work [Purcell et al., 1946], it took several decades before its potential usefulness in clinical diagnosis was fully appreciated [Damadian, 1971]. Indeed, MRI first became available in 1980 for clinical use and since then has revolutionized the field as a powerful, non-invasive and non-ionizing medical imaging technique.

The earliest brain studies were performed based on thick contiguous slices acquired

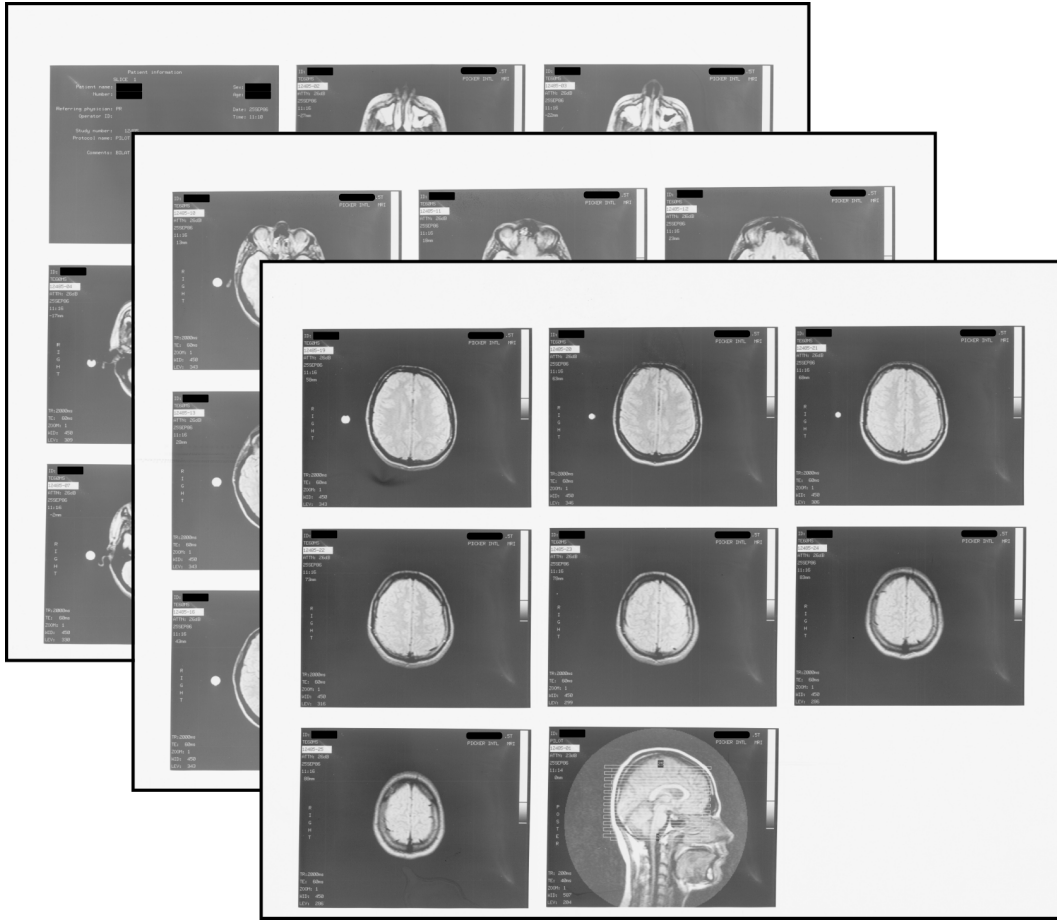


Figure 1.3.: Scanned MR Film of Multiple Sclerosis/Clinically Isolated Syndrome subject acquired at the National Hospital for Neurology and Neurosurgery in 1986 (Image adapted from [Ebner et al., 2018a]). Twenty-four proton density-like images were acquired to cover the brain and printed sequentially on three films for further visual inspection. Each slice is about 5 mm thick and was acquired at 0.5 T.

in the axial direction to cover the entire volume [Miller et al., 1988, Miller et al., 1989]. In the absence of modern standards for digital archives and visualization, the acquired scans were placed side-by-side and printed sequentially on multiple films, for further, *visual* analysis. The analysis was typically limited to measures such as lesion count and location in multiple sclerosis (MS) studies [Ormerod et al., 1987, Miller et al., 1989, Morrissey et al., 1993, O’Riordan et al., 1998]. With the introduction of the PACS (Picture Archiving and Communications System) and DICOM (Digital Imaging and COmmunication in Medicine) standards in the early 1990s, standards were created for digitally storing medical imaging information including essential

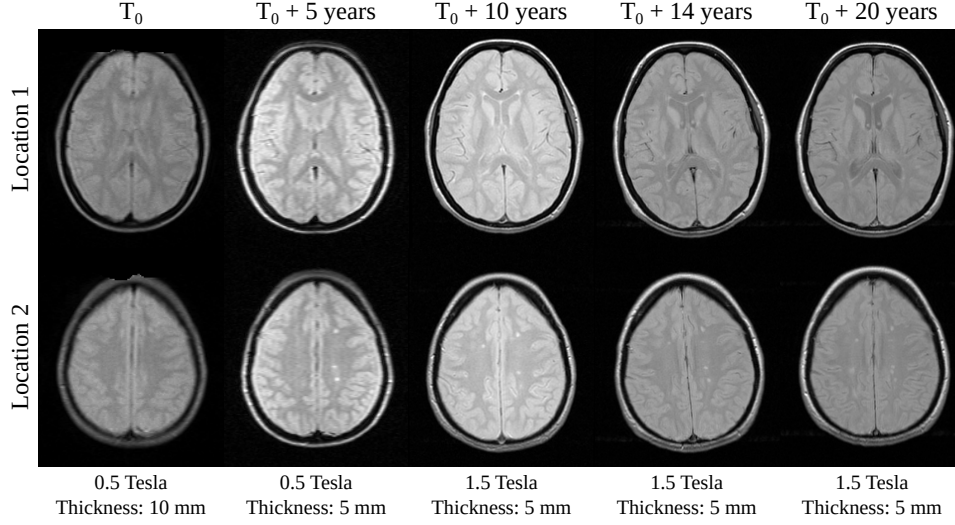


Figure 1.4.: Longitudinal proton density (PD) scans of one subject followed up since the 1980s with two spatial locations roughly aligned across acquisitions (Data courtesy of Dr. Declan Chard, UCLH NNHN). The slice thickness ranges from 10 mm in the earliest scans acquired at 0.5 T to 5 mm in more recent ones. The second row shows the evolution of the spatial distribution and size of lesions which are of particular clinical interest.

meta-data on spatial information and acquisition details. This allowed further development of clinically important biomarkers such as brain and lesion volume from longitudinal MS studies [Brex et al., 2002, De Stefano et al., 2014, Sailer et al., 1999] – information which, currently, cannot be readily extracted from scans dating back to the 1980s and early 1990s if they are only available as printed films. The original digital data is often lost or cannot be recovered due to hardware and software obsolescence issues, which also has been shown in more recent studies on brain morphometry where original MR films were digitized anew and manually processed to allow for further quantification [Ekert et al., 2016]. In other words, especially for longitudinal studies dating back to the 1980s, a decade or more of valuable image data information may not be readily accessible to modern image processing techniques, representing lost opportunities in the understanding of long-term pathological and morphological evolution.

As a motivating example, which is developed in this thesis, at the National Hospital for Neurology and Neurosurgery (NHN), Queen Square, London, a cohort

of 132 people recruited soon after a clinically isolated syndrome was diagnosed has been followed up since the 1980s with regular time points until present [Miller et al., 1988, Miller et al., 1989, Morrissey et al., 1993, O’Riordan et al., 1998, Brex et al., 2002, Fisniku et al., 2008]. A 30-year longitudinal clinical study is currently underway, which includes more than 300 image acquisitions captured on historical films. One example set of films is shown in Figure 1.3 with an additional longitudinal comparison in Figure 1.4. A framework to volumetrically reconstruct consistent 3D visualizations from printed MR films would facilitate a robust longitudinal analysis spanning more than 30 years of MRI scans. To visualize a consistent 3D geometry, such a framework needs to correctly align the sequentially printed, historical 2D slices, estimate the lost geometrical properties and dimensions, and correct for distortions from printing, storage and manual scanning of the films. Making historical scans available for modern image processing and analysis could contribute new insights and understanding of disease progression.

1.2. Super-Resolution Reconstruction: Addressing Clinical Needs of Higher Resolution in MRI

The aim of both structural and functional medical imaging is to extract information on the human body or specific organs within it [Greenspan, 2008]. Sufficiently spatial (or temporal) resolution² is vital for a better and more detailed understanding of the anatomy in addition to early detection of abnormalities and increased accuracy in the assessment of size and morphology of organs and pathologies [Greenspan, 2008]. In the context of motion, however, those needs may only be partially addressed by MRI in various applications including the mentioned areas of fetal and abdominal imaging. Ultra-fast 2D MRI, although powerful and valuable in its ability to address the problem of motion, is limited in providing detailed information on small anatomical structures due to their inherent requirement of acquiring thick anatomical slices.

Modern image processing techniques can provide an effective means to serve the

²Spatial resolution is defined as the smallest separation of two point sources necessary for the source to be resolved [Van Reeth et al., 2012].

goal of resolution augmentation post acquisition without requiring the purchase of new hardware equipment as technology evolves. Indeed, Super-Resolution (SR) methods can reconstruct high-SNR and high-resolution representations of objects that could only be acquired at low resolution because of motion, scanning time limitations or SNR considerations [Candès and Fernandez-Granda, 2014, Milanfar, 2010, Greenspan, 2008, Van Reeth et al., 2012, Plenge et al., 2012]. In several fields including cardiac MRI [Shi et al., 2013, Bhatia et al., 2014, Odille et al., 2015, Oktay et al., 2016, Bastý and Grau, 2018], thoracic MRI [Van Reeth et al., 2015], tongue [Woo et al., 2012], fetal MRI [Rousseau et al., 2010, Gholipour et al., 2010b, Kainz et al., 2015a] and, more recently, abdominal MRI [Ebner et al., 2017a] SR methods have been applied successfully in increasing the image resolution as a post-processing step.

Super-Resolution techniques can be categorized into single- and multi-image methods. Single-image SR methods³ rely on a database of previously learned relationships between low- and high-resolution images and leverage this information to enhance the resolution of a newly presented, single low-resolution image. Especially recent deep learning-based methods have been tremendously successful in single-image-based SR to create remarkable results in computer vision [Dong et al., 2016, Shi et al., 2016, Johnson et al., 2016, Lim et al., 2017] but also in medical imaging such as in cardiac MRI [Shi et al., 2013, Oktay et al., 2017, Bastý and Grau, 2018]. However, they require a very strong prior knowledge, which is not always available. This dependency on representative training data also limits the generalizability of its results, which can lead to poor performance if applied to images with different contrast, anatomies, or sizes. Moreover, motion correction is not explicitly addressed, which can constitute a critical challenge, depending on the application.

Multi-image SR methods⁴ combine information from multiple LR observations of the same object to reconstruct a single, HR image where each LR acquisition typically corresponds to an aliased output, i.e. by violating the Nyquist sampling theorem [Park et al., 2003]. The aim of SR reconstruction (SRR) methods is to combine the information present in multiple LR images to reconstruct a higher-

³also called example-based or learning-based SR methods

⁴also called reconstruction-based SR methods

resolution image [Vandewalle et al., 2010]. As pointed out in [Plenge et al., 2012], there is broad consensus that multi-image *super-resolution in MRI is not achievable in-plane, nor in true 3D acquisitions*. This has been explained by the fact that the Fourier encoding scheme excludes aliasing in the frequency and phase encoding directions [Peled and Yeshurun, 2002, Scheffler, 2002, Van Reeth et al., 2012]. The improvement in in-plane resolution, however, can lead to noise reduction and, therefore, higher SNR [Van Reeth et al., 2012, Greenspan et al., 2002]. In the literature, SRR approaches have been presented in the image domain, frequency domain, or wavelet domain [Park et al., 2003, Milanfar, 2010, Aganj et al., 2012, Gholipour et al., 2015]. Most commonly, the image domain is preferred since it offers great flexibility to model all kinds of image degradations including more sophisticated motion-models [Yang and Huang, 2010, Gholipour et al., 2015]. In order to describe the relationship between an LR observation $\mathbf{y}_k \in \mathbb{R}^{N_k}$ and the unknown, discretized HR image $\mathbf{x} \in \mathbb{R}^N$, typically the observation model [Park et al., 2003, Greenspan, 2008, Milanfar, 2010, Van Reeth et al., 2012]

$$\mathbf{y}_k = \mathbf{D}_k \mathbf{B}_k \mathbf{W}_k \mathbf{x} + \mathbf{n}_k \quad (1.1)$$

is used to describe the MR image acquisition process whereby $N_k \ll N$ for all observations $k = 1, \dots, K$ in practice. The warp operator \mathbf{W}_k describes the geometric transformation to the LR image \mathbf{y}_k from a reference position. The blurring operator \mathbf{B}_k mimics the blurring effects of the MR imaging process described by the point spread function (PSF), i.e. the impulse response of the imaging system. The downsampling operator \mathbf{D}_k describes the loss of dimensionality and \mathbf{n}_k represents the imaging noise.

Several SR algorithms have been proposed to solve this, in practice, ill-posed problem [Van Reeth et al., 2012, Milanfar, 2010, Park et al., 2003]. A common approach is the assumption of Gaussian noise \mathbf{v}_k and independent and identically distributed LR observations, which, assuming that the operators, and hence the

motion, are known, lead to the maximum a posteriori (MAP) formulation

$$\min_{\mathbf{x}} \left(\sum_{k=1}^K \|\mathbf{y}_k - \mathbf{D}_k \mathbf{B}_k \mathbf{W}_k \mathbf{x}\|_{\ell^2}^2 + \alpha \Phi(\mathbf{x}) \right) \quad (1.2)$$

with regularization parameter $\alpha > 0$ and prior term $\Phi(\mathbf{x})$ [Yang and Huang, 2010]. Examples include Tikhonov regularization with $\Phi(\mathbf{x}) = \|\mathbf{Q}\mathbf{x}\|_{\ell^2}^2$ where \mathbf{Q} represents a high-pass filter such as the first derivative [Tanaka and Okutomi, 2010], and total variation regularization [Yang and Huang, 2010, Basti et al., 2017].

Several challenges are associated with solving (1.2) in practice. Firstly, one relies on the assumption that all operators in (1.1) are known or can be estimated reliably. Given the nature of the involved operators, this can result in large linear systems involving large matrices that can incur a substantial memory and computational cost. In its classical theory [Park et al., 2003, Yang and Huang, 2010, Hardie, 2010], SR is possible only when the LR observations are subsampled, i.e. aliased, as well as *shifted with subvoxel precision*. This guarantees that each LR observation contains new image information that can be exploited to generate an HR image [Park et al., 2003, Hardie, 2010]. Consequently, SR can only work accurately when there is very precise motion estimation with subvoxel accuracy of all LR observations for the recovery of subvoxel details. The associated registration problem is critical but well known to be ill-posed, especially in the context of SR where the associated observations are of low resolution [Yang and Huang, 2010]. Moreover, without knowing the PSF it is in principle impossible to register precisely images blurred by motion [Sorel et al., 2010]. In practice, however, the blurring operator is typically set to be known and spatially invariant to keep the problem tractable. Some SRR methods like non-local means or probabilistic motion modeling aim to bypass explicit motion estimation at the expense of increased computational complexity [Protter et al., 2009, Takeda et al., 2009, Protter and Elad, 2010]. Other approaches consider the registration problem separate from HR image estimation. However, as highlighted in [Yang and Huang, 2010], *LR image registration and the HR image estimation are actually interdependent. On one hand, accurate subpixel motion estimation benefits HR image*

estimation. On the other hand, high quality HR image can facilitate accurate motion estimation.

As of today, little is known about the optimal number of input LR images for optimal SRR results and theoretical predictions have been made only for very simple motion scenarios like translation [Baker and Kanade, 2002, Lin and Shum, 2004]. For general, more realistic, situations, the optimal number of input LR images is not known and will be application dependent [Poot et al., 2010, Rousseau et al., 2010, Van Reeth et al., 2012]. In particular, the number of LR images will depend on the magnification factor, i.e. the ratio between the resolution of the reconstructed HR image of the SR algorithm and the input LR images, but no consensus or shared best practices are available. Therefore, the SRR problem (1.1) can be severely ill-posed, especially when the magnification factor is large, e.g. greater than 2 [Yang and Huang, 2010]. In particular, for large enough magnification factors any smoothness prior leads to overly smooth results with very little high-frequency content regardless of the number of LR input images used [Baker and Kanade, 2002].

1.3. Thesis Contributions

The overall goal of this thesis is to provide robust and versatile volumetric reconstruction approaches that can be applied in a variety of clinical applications where motion patterns are inherently different, but the use of thick-slice MR data is current clinical practice. Several challenges needed to be overcome in this context which mainly relate to two key problems mentioned in the previous section: *Reliable and accurate motion correction* of individual, low-resolution 2D slices and *robust volumetric reconstruction* to achieve a high-fidelity, high-resolution 3D visualization from the motion-corrected, and potentially sparse, scattered data.

1.3.1. Super-Resolution Reconstruction of Fetal Brain MRI

For fetal brain MRI reconstruction, multiple high-resolution reconstruction frameworks have been presented in the literature already [Rousseau et al., 2006, Jiang et al., 2007, Kim et al., 2010b, Gholipour et al., 2010b, Kuklisova-Murgasova et al.,

2012, Kainz et al., 2015b]. Although the proposed approaches are shown to be useful in practice, the unpredictable and large fetal motion and its associated challenges require particularly robust motion-correction and volumetric reconstruction frameworks to allow for potential clinical translation. In particular, previously proposed robust super-resolution approaches rely on multiple hyperparameters to be tuned in order to achieve optimal reconstruction outcomes. Moreover, the proposed approaches require time-consuming optimization methods due to their resulting non-convex minimization problem formulation [Gholipour et al., 2010b, Kuklisova-Murgasova et al., 2012]. Obtaining the HR fetal brain reconstructions in the standard anatomical planes can facilitate brain studies and is typically favoured for clinical assessment by clinicians. However, currently existing methods for automatic visualization in the standard anatomical planes may work for mildly pathological cases only [Tournier et al., 2017]. Moreover, recently proposed automatic brain segmentations approaches [Anquez et al., 2009, Taleb et al., 2013, Keraudren et al., 2014, Rajchl et al., 2016, Salehi et al., 2018] unlock the potential use of automatic fetal brain reconstruction studies in clinical practice. Yet, to date, no reconstruction framework has been tested as part of a fully-automated brain MRI reconstruction pipeline requiring no manual intervention. This thesis aims to address these shortcomings and presents the following contributions [Ebner et al., 2018c, Ebner et al., 2019c]:

- An alternative outlier-robust volumetric reconstruction framework for isotropic, high-resolution 3D reconstruction from motion-corrupted, low-resolution 2D slices.
- A simple, yet demonstrably effective, complete outlier-rejection method for robust super-resolution that relies on a *single* hyperparameter only and retains a *linear least-squares formulation* which can be solved efficiently.
- A *fast and robust template-space alignment* approach for automatic high-resolution visualization in the standard anatomical planes *suitable for pathological brains*.
- A motion-correction method based on Gaussian process regularization that encourages the consistency of spatial transformations with the overall objective

of achieving more robust slice motion-correction estimates.

- Demonstration of the clinical potential of the presented reconstruction framework as part of a fully automated localization, segmentation and super-resolution reconstruction framework for fetal brain MRI that *requires no manual intervention*.

1.3.2. Super-Resolution Reconstruction of Upper Abdominal MRI

In the context of upper abdominal MRI, no algorithm has been presented to reconstruct an isotropic, high-resolution 3D visualization to define upper gastrointestinal (GI) tract and biliary anatomy from low-resolution SST2W sequences. Despite super-resolution being well suited in principle, its application is particularly challenging in the context of MR cholangio-pancreatography (MRCP) studies due to the sparse data in clinical acquisition protocols (only two SST2W scans; one axial and one coronal acquisition) and the necessity of correcting for complex, non-rigidly deforming anatomy. However, as opposed to fetal MRI, abdominal MRI is characterized by less substantial motion and allows for better-controlled experiments, which can enable valuable validation studies of motion-correction and volumetric reconstruction algorithms. Moreover, current reconstruction frameworks presented for the high-resolution reconstruction of SST2W data do not take into account the image formation process for registration/motion-correction. This is particularly problematic since neglecting the point spread function (PSF) during resampling introduces aliasing and subsequently results in additional loss of information [Cardoso et al., 2015, Chacko et al., 2015]. In this thesis, the following contributions are presented:

- Introduction of a super-resolution reconstruction framework to reconstruct an isotropic, high-resolution 3D visualization of upper GI anatomy specifically tailored to the sparsely available data in clinical MRCP studies [Ebner et al., 2017a].
- Introduction of PSF-aware registration that takes into consideration the physical image formation process during motion correction for MRCP sequences [Ebner

et al., 2017a].

- Design and execution of a pilot study involving the acquisition of multiplanar SST2W stacks of the upper abdomen and the brain ('quasi-static' control data, to remove the effect of upper abdominal motion artefact) for eight healthy volunteers with the overall objectives of [Ebner et al., 2019a, Chouhan et al., 2019, Ebner et al., 2019b]
 1. optimising source image acquisition protocols by establishing the ideal number and orientation of SST2W series for MRCP SRR generation
 2. optimising post-processing protocols by defining the most suitable approach to registration/motion correction for SRR in the upper abdomen
 3. validating the overall potential of upper abdominal SRR using expert-readers

1.3.3. Volumetric Reconstruction from Printed Films

Historical, printed brain MR films are not readily accessible to modern image processing techniques. A volumetric reconstruction needs to account for various kinds of image degradations and geometrical distortions present in printed films. This thesis presents the following contributions [Ebner et al., 2018a]:

- A volumetric reconstruction framework to recover geometrically consistent, volumetric images from printed films in MR neuroimaging.
- The reconstruction framework is based on a joint slice-to-volume affine registration with inter-slice 2D transformation regularisation and affine slice-intensity correction whereby the missing meta-data information is contributed by a longitudinal scan of the same subject.
- A final isotropic total variation in-plane deconvolution technique to revitalise the visual appearance of the reconstructed stack of historical slices.

1.3.4. Forward-Backward Splitting in Deformable Image Registration

Efficient non-linear image registration implementations are key for many biomedical imaging applications. Recently proposed forward-backward splitting approaches represent promising numerical schemes that, although particularly relevant for medical image registration, have not been widely used in this context. Associated contributions in this thesis are [Ebner et al., 2018b]:

- Introduction of recently proposed Forward-Backward Splitting (FBS) methods to deformable image registration.
- Highlighting that FBS is guaranteed to converge to a critical point for cost functions of the form $f + g$ consisting of a smooth (possibly non-convex) function f and a convex (possibly non-smooth) function g . In the context of (non-linear) image registration, f will typically correspond to a (differentiable) similarity measure and g to a (convex) regularizer.
- Showcasing the advantage of FBS to efficiently solve for various kinds of cost functions vital for medical image registration by performing two simple iterative steps – a *forward* and a *backward* step.
- Illustration that Tikhonov regularization breaks down to simple B-Spline filtering in the backward step.
- Demonstration of the versatility of FBS by encoding spatial transformation as displacement fields or free-form B-Spline deformations.
- Comparison of two FBS variants, namely FISTA [Goldstein et al., 2014] and iPiano [Ochs et al., 2014], against the classical demons algorithm, the recently proposed inertial demons algorithm [Santos-Ribeiro et al., 2016] and the conjugate gradient method within NIFTYREG [Modat et al., 2010].
- Numerical experiments performed on both synthetic and clinical data show the advantage of FBS in image registration in terms of both convergence and accuracy.

1.3.5. Publicly Available Open-Source Implementations

To facilitate collaboration both within and outside of academia, the software implementations of the developed concepts have been made publicly available as open-source packages:

- NIFTYMIC⁵: A generic toolkit for motion correction and volumetric image reconstruction of 2D ultra-fast MRI. A robust super-resolution reconstruction implementation is provided for a variety of robust data loss function and regularizer options.
- ITK_NIFTYMIC⁶: An extension to the Insight Segmentation and Registration Toolkit⁷ (ITK) framework to enable PSF-aware registration and volumetric reconstruction.
- VOLUMETRICRECONSTRUCTIONFROMPRINTEDFILMS⁸: A toolkit to reconstruct a volumetric representation from printed brain MR films.
- NSOL⁹: A numerical solver library for denoising and deconvolution applications using Alternating Direction Method of Multipliers (ADMM) and Primal-Dual methods.
- SIMPLEREG¹⁰: A toolkit providing conversion interfaces between commonly used registration platforms including NIFTYREG¹¹, FLIRT¹² and ITK-based frameworks in addition to extended registration functionalities.

1.4. Thesis Organisation

This thesis is structured in six chapters. In the **current chapter**, the challenges and limitations of MR imaging in the context of motion were pointed out. Ultra-fast 2D

⁵<https://github.com/gift-surg/NiftyMIC>

⁶https://github.com/gift-surg/ITK_NiftyMIC

⁷<https://itk.org>

⁸<https://github.com/gift-surg/VolumetricReconstructionFromPrintedFilms>

⁹<https://github.com/gift-surg/NSoL>

¹⁰<https://github.com/gift-surg/SimpleReg>

¹¹<https://github.com/KCL-BMEIS/niftyreg>

¹²<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT>

MRI was described as a means to mitigate this limitation by acquiring an image stack of thick 2D slices spanning the anatomy of interest for further clinical assessment. Two important clinical applications of fetal and abdominal MRI were presented, for which ultra-fast 2D MRI plays a major role in current clinical practice. The concept of Super-Resolution and its challenges were presented to address the clinical need of higher-resolution visualization to allow for more accurate and reliable clinical diagnosis and facilitated disease management. Motion correction was identified as one of the key challenges to be addressed to reconstruct a high-resolution volume of diagnostic value by Super-Resolution algorithms. The inherently different nature of unpredictable, large motion (fetal MRI), pseudo-periodic, breathing and bowel motion (abdominal MRI) and random motion (historical MR films) was discussed to enumerate a variety of potential clinical applications that could benefit from robust and versatile motion-correction frameworks.

The **second chapter** (Super-Resolution Reconstruction of Fetal Brain MRI), starts with a review on Super-Resolution Reconstruction frameworks developed for fetal brain MRI. Given the unpredictability of fetal motion, the application of Super-Resolution Reconstruction algorithms on, potentially, severely motion-corrupted, LR stacks of slices is widely regarded as a highly challenging problem. A framework for automatic localization, segmentation and Super-Resolution Reconstruction of fetal brain MRI is presented. The proposed pipeline produces automatic reconstructions that are comparable to manual segmentation-based reconstructions, therefore, effectively eliminating the need for manual intervention.

Compared to fetal MRI, abdominal MRI is characterized by less severe motion where, typically, additional types of MR acquisitions such as a heavily T2-weighted volume are acquired during MRCP studies. In the **third chapter** (Super-Resolution Reconstruction of Upper Abdominal MRI), specifically tailored MRCP SRR approaches are presented whereby controlled abdominal and 'quasi-static' experiments are performed for validation purposes. Additionally, it introduces the concept of PSF-aware slice-to-volume registration in the context of ultra-fast 2D MRI.

The **fourth chapter** (Volumetric Reconstruction from Printed Films: Enabling 30

Year Longitudinal Analysis in MR Neuroimaging) is dedicated to the volumetric reconstruction from printed, historical MR films. The presented reconstruction framework is based on semi-automatic slice extraction, automated in-plane 2D affine registration with inter-slice transformation regularization and slice intensity correction followed by isotropic total variation reconstruction. This framework will be used to facilitate the robust analysis of a uniquely long-term multiple sclerosis study spanning 30 years of MRI scans of people followed up after a clinically isolated syndrome.

The **fifth chapter** (Forward-Backward Splitting in Deformable Image Registration: A Demons Approach) focuses on a means to efficiently solve non-linear image registration problems. Taking advantage of the rich mathematical body developed for Forward-Backward Splitting (FBS) methods in recent years, it shows that registration problems with any combination of differentiable (not necessarily convex) similarity measure and convex (not necessarily smooth) regularization with a tractable proximal operator can be solved efficiently. Moreover, it describes how these insights could be useful for achieving more robust slice-to-volume registration estimates for the presented reconstruction frameworks.

Finally, the **sixth chapter** (Conclusions and Future Work) concludes the thesis and outlines potential future research directions.

CHAPTER 2.

Super-Resolution Reconstruction of Fetal Brain MRI

Table of Contents

2.1. Introduction	59
2.2. An Automated Framework for Localization, Segmentation and Super-Resolution Reconstruction of Fetal Brain MRI	68
2.2.1. Localization based on Coarse Segmentation	69
2.2.2. Fine Segmentation	71
2.2.3. Robust HR Volume Reconstruction Framework	73
2.3. Experiments and Results	79
2.3.1. Data	79
2.3.2. Implementation Details	80
2.3.3. Localization Results	83
2.3.4. Segmentation Results	84
2.3.5. Outlier-Robust SRR Results	86
2.4. Discussion	95

Foreword. This chapter is adapted and contains content from the work presented in

- **Ebner, M.**, Wang, G., Li, W., Aertsen, M., Patel, P. A., Aughwane, R., Melbourne, A., Doel, T., David, A. L., Deprest, J., Ourselin, S., and Vercauteren, T. (2018c). An Automated Localization, Segmentation and Reconstruction Framework for Fetal Brain MRI. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2018*, pages 313–320. Springer
- **Ebner, M.**, Wang, G., Li, W., Aertsen, M., Patel, P. A., Aughwane, R., Melbourne, A., Doel, T., Dymarkowski, S., De Coppi, P., David, A. L., Deprest, J., Ourselin, S., and Vercauteren, T. (2019c). An Automated Framework for Localization, Segmentation and Super-Resolution Reconstruction of Fetal Brain MRI. *Under Submission (NeuroImage)*

My specific contributions presented in this chapter¹ are:

- A publicly available² outlier-robust volumetric reconstruction framework for isotropic, high-resolution 3D reconstruction from motion-corrupted, low-resolution 2D slices for fetal brain MRI.
- My methodological contributions of this framework include:
 - An effective complete outlier-rejection method for robust super-resolution reconstruction that relies on a single hyperparameter only and retains a linear least-squares formulation.
 - A fast and robust template-space alignment approach for automatic HR visualization in the standard anatomical planes suitable for pathological brains.
 - A robust motion-correction method based on Gaussian process regularization that encourages the consistency of spatial transformations within each sub-stack of an interleaved stack acquisition³.
- The reconstruction results of the proposed method compare favourably with those obtained by manual, labour-intensive brain segmentations, which unlocks the potential use of automatic fetal brain reconstruction studies in clinical practice.

¹The work in this chapter associated with the automatic fetal brain localization and segmentation steps was contributed by Guotai Wang. It was made available in this thesis to provide context and ease readability of the overall framework.

²<https://github.com/gift-surg/NiftyMIC>

³Unpublished work and an extension to [Ebner et al., 2018c, Ebner et al., 2019c].

A follow-up work is currently in preparation where the fully automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI presented in this chapter is applied to automatically reconstruct more than 300 fetal MRI examinations. In particular, this framework will be used to help quantify the impact of spina bifida surgical closure [Sacco et al., 2018].

2.1. Introduction

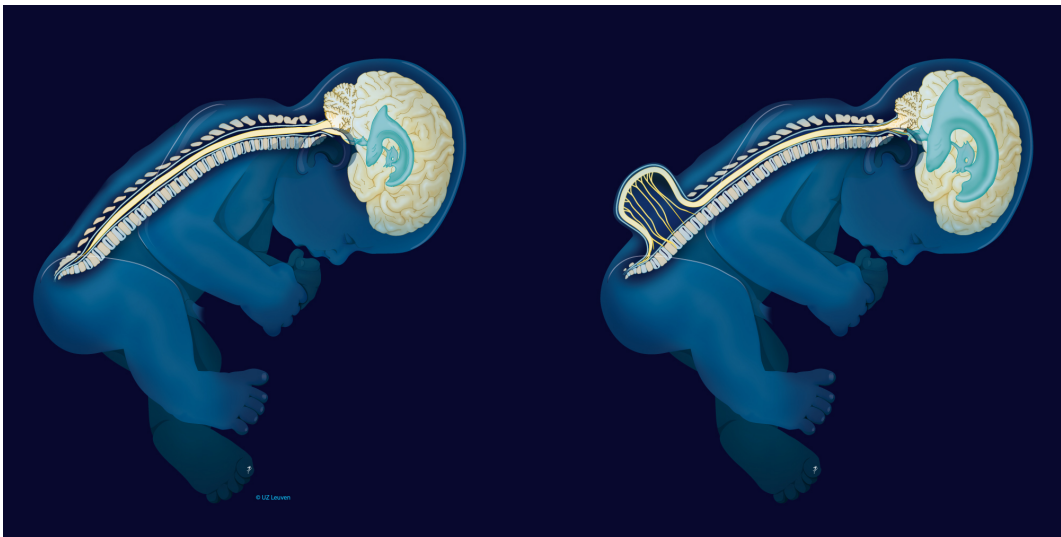


Figure 2.1.: Chiari II malformation with ventriculomegaly. Courtesy of Prof. Anna David, UCLH. ©University College London Hospitals NHS Foundation Trust.

As briefly discussed in Chapter 1, Fetal Magnetic Resonance Imaging (MRI) has become increasingly important in prenatal diagnosis as a complementary tool to ultrasound, for its advantages in demonstrating pathologies in soft tissues, that may not be apparent or cannot be accurately characterized on prenatal ultrasonography. To mitigate the effect of fetal (and maternal) motion, fast single-shot T2-weighted (SST2W) sequences are used to acquire thick, low-resolution (LR) stacks of 2D slices that can largely freeze in-plane motion [Saleem, 2014, Gholipour et al., 2014]. With motion commonly occurring in between slice acquisitions, this generally results in motion-corrupted stacks of slices in multiple orientations with poor 3D image integrity and resolution. In order to assess and quantify fetal brain development and

pathology, it is highly desirable to reconstruct a single isotropic, high-resolution (HR) volume of the fetal brain in standard anatomical planes from multiple LR stacks acquired in different views. One indication for fetal MRI is spina bifida, where MRI plays a role in characterizing the spinal lesion as well as the associated brain changes [Aertsen et al., 2019]. In open spina bifida (myelomeningocele and myeloschisis), a fault in the development of the spinal cord and surrounding vertebrae leaves a gap in the spine, allowing the spinal cord and nerve tissue to bulge through a defect on the baby’s back. Because of a suction gradient by leakage of cerebrospinal fluid at the lesion, the hind brain descends through the base of the skull where the spinal cord exits (a condition termed a Chiari II malformation, Figure 2.1). This may be associated with excessive accumulation of fluid in the brain ventricles (ventriculomegaly). In these cases, HR 3D reconstructions would aid more accurate measurements, currently performed on LR 2D stacks [Aertsen et al., 2019], and help characterize associated brain changes, ruling out those that are prognostically important.

Currently existing reconstruction toolkits generally rely on an approach that iteratively operates motion correction and super-resolution reconstruction (SRR) [Rousseau et al., 2006, Gholipour et al., 2010b, Kuklisova-Murgasova et al., 2012, Kainz et al., 2015b]. Since the position and orientation of the fetal brain vary significantly between different patients in relation to maternal structures, localizing the fetal brain and obtaining a segmented mask to exclude the surrounding tissues is crucial to achieve accurate motion correction outcomes. Current motion-correction approaches typically employ rigid registration with the assumption that the brain has rigid and surrounding tissues non-rigid motion patterns. Thus, localization/segmentation can help to clearly delineate the brain region so that rigid motion correction becomes meaningful. At present, this usually requires manual localization of the fetal brain and uses manual or semi-automatic methods to obtain fetal brain masks, which is laborious and time consuming. The approach of [Alansary et al., 2017] avoids this problem by directly reconstructing the whole uterus by splitting each slice into patches for motion correction. However, apart from the increased computational

requirements and reduced robustness, this leads to non-rigid motion correction and, thus, suboptimal outcomes of rigidly moving regions such as the fetal brain. Moreover, it is highly desirable to obtain the HR fetal brain visualization in the standard anatomical planes for clinical assessment. For this purpose, previously proposed automatic methods, such as presented in [Tourbier et al., 2017] or [Gholipour et al., 2017], rely on gestational age-matching between *normal* brains for the template brain selection. However, especially for pathological brains such as encountered with spina bifida [Ovaere et al., 2015, Aertsen et al., 2019], severe morphological and brain volume changes require a more robust template space alignment approach. Hence, we hypothesize that a fully automatic reconstruction pipeline based on *automatic* fetal brain localization, segmentation and *robust* reconstruction and template-space alignment steps is favorable to achieve efficient and accurate fetal brain reconstructions for potential clinical translation.

Several studies have reported automatic fetal brain localization and segmentation. They are based on template-registration [Taleb et al., 2013], template-matching [Anquez et al., 2009] or machine learning with hand-crafted features of local patches [Keraudren et al., 2014]. Whereas template-based methods are not robust to pathologies and large motions, traditional learning methods with local patches suffer from inefficiency at inference time. Convolutional neural networks (CNNs) have been used for this task [Rajchl et al., 2016, Salehi et al., 2018] for better performance. However, since the fetal brain is relatively small in a whole slice and these CNNs work on the whole slice, they can easily generate false positives [Salehi et al., 2018]. In addition, their utility for automatic fetal brain reconstruction has yet to be demonstrated.

With the localized or segmented fetal brain masks, multiple motion-corrupted LR stacks can be reconstructed into a single HR volume. There are two main challenges in the HR volume reconstruction step. First, the inter-slice motion can lead to inconsistent appearance in neighboring slices. This is mainly due to the fact that the SST2W sequence acquires fetal MR images in an interleaved fashion to reduce the scan time and avoid slice cross-talk artifacts [Gholipour et al., 2014]. An M -interleaved scanning leads to M sub-stacks that are temporally sequential but

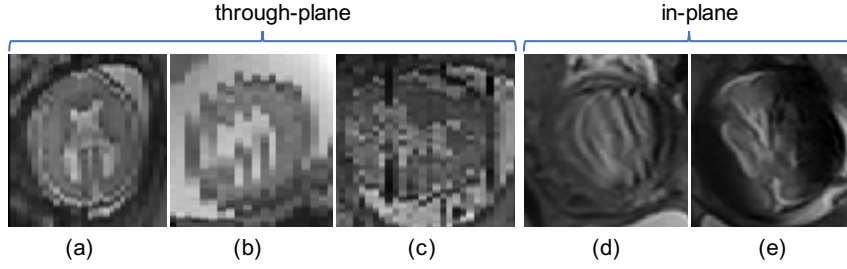


Figure 2.2.: Three example stacks of fetal brain MRI. (a) has a consistent appearance and small inter-slice motion. (b) has an inconsistent motion between two interleaved sub-stacks. (c) has two outlier slices shown in (d) and (e). The gestational age is 24 weeks for the fetuses shown in (a) and (b), and 29 weeks for the fetus shown in (c)-(e).

spatially interleaved, where M is usually set as 2 or 3. The motion pattern within each sub-stack is relatively consistent and smooth while that between sub-stacks can be inconsistent, as shown in Figure 2.2(b). Moreover, motion during image acquisition can lead to various types of artefacts such as in-plane image blur, slice crosstalk and spin-history artefacts that can considerably affect the image quality of individual slices [Gholipour et al., 2014] as visualized in Figure 2.2(c)-(e). For motion correction, previously described methods typically rely on a hierarchical approach which starts from entire sub-stacks and subdividing them until each slice is dealt with independently [Rousseau et al., 2006, Gholipour et al., 2010b]. However, due to the unpredictable nature of fetal motion, the generic subdividing might lead to a suboptimal grouping of slices for motion correction for which the approaches in [Kuklisova-Murgasova et al., 2012, Kainz et al., 2015b] present rigid slice-to-volume registrations directly after a first volumetric alignment. Second, robustness against *outlier* slices characterized by either misregistration or image artefacts is key for a high-fidelity HR reconstruction [Gholipour et al., 2010b, Kuklisova-Murgasova et al., 2012]. However, no complete outlier slice rejection is achieved in [Gholipour et al., 2010b] and the method of [Kuklisova-Murgasova et al., 2012, Kainz et al., 2015b] relies on multiple hyperparameters to be tuned in order to achieve optimal reconstructions while both require time-consuming optimization methods due to their resulting non-convex problem formulation.

We propose a novel CNN-based fetal brain segmentation method in a coarse-to-fine fashion to reduce false positives and applied the proposed framework for HR volume reconstruction of MR images of fetuses with spina bifida. Additionally, we introduce an approach for effective complete rejection of detected outlier slices as an alternative to the methods in [Gholipour et al., 2010b, Kuklisova-Murgasova et al., 2012], that relies on a single hyperparameter only and retains a linear least-squares formulation which can be solved efficiently. In this chapter, we give a detailed description of the framework including four novel aspects:

1. we investigate a novel robust motion correction method that encourages consistency of spatial transformations within each sub-stack of an interleaved stack acquisition
2. we propose a fast template space alignment method that is robust to large brain morphology changes such as encountered in spina bifida
3. we show the superiority of our automatic localization, segmentation and reconstruction methods by comparing them with different variants and state-of-the-arts
4. we validate the proposed framework with a large dataset of images from both normal fetuses and fetuses with spina bifida [Aertsen et al., 2019]

Experimental results show that our framework can achieve comparable reconstruction output to that of manual segmentation-based reconstruction, and it outperforms existing fetal brain extraction and reconstruction methods on different cohorts of fetuses.

Related Works

Fetal Brain Localization and Segmentation

Extracting fetal brain from fetal MRI usually serves as a prerequisite step for HR volume reconstruction of the fetal brain. [Anquez et al., 2009] proposed an automated

fetal brain segmentation method by localizing the fetal eyes and then segmenting the neighboring skull bone content, which can lead to a poor performance when the inter-slice motion is large. [Taleb et al., 2013] used a template-based method to generate fetal brain masks. It obtains a region of interest (ROI) based on the intersection of multiple scans of the same patient, and then registers the ROI to an age-specific template. [Tourbier et al., 2017] used template-to-slice block matching and deformable slice-to-template registration for automatic fetal brain localization and segmentation. It achieved good performance at the cost of a very long computational time up to several hours. [Kainz et al., 2014a] proposed to localize the fetal brain by voxel classification using rotation invariant volume descriptors. [Keraudren et al., 2013] used bundled Scale-Invariant Feature Transform (SIFT) features to fit a 3D bounding box of the fetal brain where prior knowledge of the fetal brain development was used to define size and shape constraints for robust localization. [Keraudren et al., 2014] extended that method for fetal brain segmentation with image-specific online learning based on Random Forests. It is limited by hand-crafted features and inefficiency during inference.

Recently, deep learning with CNNs has been employed for fetal brain segmentation from fetal MRI. [Rajchl et al., 2016] used a fully convolutional neural network (FCN) for fetal brain segmentation under distributed weak supervision. [Salehi et al., 2018] used a 2D U-Net [Ronneberger et al., 2015] for slice-by-slice fetal brain segmentation. These methods with CNNs predict the segmentation result directly without a localization step and, compared with previous methods, are more efficient at test time. However, they can easily cause false positives and show poor performance for challenging cases, and their utility has not been demonstrated in the context of automatic fetal brain reconstruction.

Fetal Brain Reconstruction

Slice-based motion correction techniques have been successfully applied in various fields of medical imaging [Ferrante and Paragios, 2017] and can estimate the correct inter-spatial relationship of multiple slices to reconstruct a consistent volumetric

representation. To date, slice-based volumetric reconstruction methods have been applied to fetal brain [Rousseau et al., 2006, Jiang et al., 2007, Kim et al., 2010b, Gholipour et al., 2010b, Kuklisova-Murgasova et al., 2012, Kainz et al., 2015b], fetal thorax [Kainz et al., 2014b], fetal trachea [Ebner et al., 2016], the whole uterus including the human and the placenta [Alansary et al., 2017, Kainz et al., 2015a], and the fetal heart [Lloyd et al., 2019] in the context of fetal MRI. An extended review of previously proposed SRR frameworks for fetal brain MRI is provided in the following.

[Rousseau et al., 2005, Rousseau et al., 2006] proposed a slice-to-volume registration (SVR) method for fetal brain reconstruction based on semi-automatic segmentation results. It consisted of three steps: motion correction, volume reconstruction and contrast correction. In the motion correction step, each LR stack is globally aligned first followed by a hierarchical slice package motion correction approach based on the temporal slice interleave. Iterative reconstructions are used as reference for motion correction which were obtained by using scattered interpolation with a narrow Gaussian kernel as the point spread function (PSF). A contrast correction step is used to correct the local relative intensity distortion between the LR stacks. [Jiang et al., 2006, Jiang et al., 2007] used a similar approach but used multilevel scattered cubic B-spline interpolation for the reconstruction task instead which resulted in isotropic HR volumes with reduced blurring. Additionally, they investigated using multiple LR stacks acquired only in the transverse direction with overlapping slices instead of orthogonally acquired LR image stacks in all three anatomical planes. Apart from the non-compatibility of this protocol with standard clinical practice, where typically orthogonal slices are acquired for clinical assessment, the single-orientation input data also led to worse through-plane quality in the obtained HR volumes compared to using multi-orientation input. [Kim et al., 2008, Kim et al., 2010b] proposed a reconstruction-free registration approach that relies on a slice intersection motion correction (SIMC) method that directly co-aligns multiple stacks of 2D slices which was followed by a single Gaussian-weighted averaging step for the volumetric reconstruction. Subsequent works incorporated a non-iterative relative

bias field correction prior to SIMC to attenuate bias field inhomogeneities across slices [Kim et al., 2011c] and volumetric reconstructions with improved noise reduction by using structure tensor weighted kernel regression [Kim et al., 2011a]. Their resulting software tool is called the “SLIce MRI Motion Estimation and Reconstruction” (SLIMMER)⁴ [Kim et al., 2011b]. [Gholipour and Warfield, 2009, Gholipour et al., 2010a, Gholipour et al., 2010b] formulated the volumetric reconstruction step as a super-resolution reconstruction problem (1.2) that allowed a minimum error representation of the obtained HR volume, whereby the slice acquisition model (1.1) was used. For increased robustness, it used a robust M-estimation formulation in (1.2) that minimizes a Huber’s error function to *reduce* the influence of potential outliers. [Kuklisova-Murgasova et al., 2012] built on the idea of robust SRR and proposed a reconstruction method with *complete* outlier rejection that can entirely exclude identified misregistered slices or corrupted voxels using expectation-maximization (EM)-based robust statistics. Intensity matching was used to compensate inconsistent scaling factors and bias fields of acquired slices as motivated by [Ashburner and Friston, 2005]. Additionally, an edge-preserving prior function [Charbonnier et al., 1997] was proposed as a regularization term for the SRR formulation (1.2). [Rousseau et al., 2010] investigated experimentally the question on the ideal number of LR stacks in fetal MRI required for super-resolution to trade-off reconstruction quality and acquisition time. Three orthogonal stacks, one in each anatomical direction, were identified as the minimum requirement for fetal brain SRR under optimal conditions as tested on synthetic, noise-free data. However, experiments on the optimal number for fetal brain SRR remained inconclusive but they argued that denoising can substantially decrease the number of required LR images. As a follow-up work, the “Baby Brain Toolkit” (BTK)⁵ was published as open-source reconstruction framework for fetal brain MRI [Rousseau et al., 2013] by using blockwise non-local means (NLM) denoising of [Buades et al., 2005, Coupé et al., 2008] and a non-local regularization term for the SRR formulation (1.2) based on [Manjón et al., 2010].

⁴<http://depts.washington.edu/bicg/software/SLIMMER/?f=slimmer.html>,
Last accessed: May 4, 2019 (SLIMMER download was not available)

⁵<https://github.com/rousseau/fbrain>

[Tourbier et al., 2014, Tourbier et al., 2015] proposed a total variation (TV) regularization for the SR step as part of BTK, which was solved using a primal-dual algorithm [Chambolle and Pock, 2011]. [Kainz et al., 2015b] developed a GPU-accelerated implementation⁶ of [Kuklisova-Murgasova et al., 2012], and proposed to automatically select the stack with least motion as the reference stack. Additionally, the PSF was modeled as a sinc-function in the in-plane directions for the SRR step, therefore, being closer to the theoretically assumed PSF [Liang and Lauterbur, 2000, Jiang et al., 2007]. Later, [Kainz et al., 2015a, Alansary et al., 2017] proposed a patch-to-volume registration (PVR) framework to reconstruct larger field-of-views beyond the fetal brain by breaking a slice into smaller, overlapping patches at the cost of increased computational requirements. This poly-rigid reconstruction framework approximates more complex, non-rigid motion and was applied to reconstruct flexible organs including the placenta [Kainz et al., 2016, Miao et al., 2017] and the fetal heart [Lloyd et al., 2019]. [Tourbier et al., 2017] proposed a fully automated reconstruction pipeline including template-space alignment step for the HR visualization of the fetal brain in the standard anatomical planes but presented gestational age-matching to select the template from the normal brain atlas which may work for mildly pathological cases only. Apart from the SIMC approach, all other presented approaches so far relied on an iterative two-step motion-correction/volumetric reconstruction approach. However, despite shown to be effective in practice, a separation into two steps does not guarantee optimality in the estimated volumetric reconstruction as defined by (1.1). [Fogtmann et al., 2012] aimed to address this shortcoming by framing the motion-correction and volumetric reconstruction problem as a single, unified formulation. Additionally, they proposed a generic motion prior to penalize differences between slice-motion correction parameters of temporal neighboring slices. However, the practical implementation of the proposed UF followed again an iterative paradigm that separates the steps slice transformation and volumetric reconstruction steps. [McDonagh et al., 2017] proposed a context-sensitive upsampling method based on CNNs to improve the resolution of each LR stack, and then

⁶<https://github.com/bkainz/fetalReconstruction>

used the upsampled LR stacks as the inputs of an SVR-based 3D reconstruction method. [Hou et al., 2017, Hou et al., 2018] utilized CNNs to predict the initial transformation parameters of SVR in the motion correction step to achieve more robust initialisations for the slice-to-volume registration step.

Overall, the rich body of research presented in the literature illustrates the multitude of challenges encountered in fetal MRI associated with the critical steps of automatic brain localization, segmentation and robust high-resolution reconstruction in the standard anatomical planes. Nevertheless, to the best of our knowledge, the existing methods have not yet fully solved all these issues vital for a robust and fully-automated reconstruction framework that may allow clinical translation.

2.2. An Automated Framework for Localization, Segmentation and Super-Resolution Reconstruction of Fetal Brain MRI

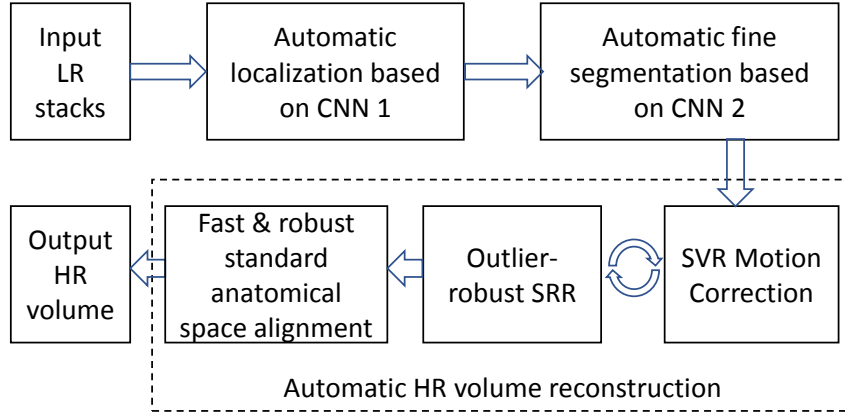


Figure 2.3.: The proposed fully automatic framework for fetal brain reconstruction. The automatic localization, segmentation and reconstruction parts are detailed in Figure 2.4, Figure 2.5 and Figure 2.6 respectively.

An overview of our proposed fully automatic framework for fetal brain reconstruction is depicted in Figure 2.3. We first use a CNN to automatically localize the fetal brain region in each input LR stack and obtain a 3D bounding box of the fetal brain. Within the bounding box, we use another CNN to automatically generate a fine seg-

mentation of the fetal brain. The automatic HR volume reconstruction stage includes the two-step iterative SVR and outlier-robust SRR step followed by a fast and robust standard anatomical template space alignment step. For the outlier-robust SRR, we propose a novel outlier rejection (OR) method by defining a similarity measurement to remove outlier slices and frame the SRR problem as a linear least-squares formulation that can be solved efficiently. For robust SVR, we investigate Gaussian Process Regularization (GPR) to encourage spatial consistency of the transformations of slices that are within the same sub-stack of an interleaved acquisition. For the template registration step, we propose a rigid registration approach based on symmetric block-matching between the SRR and a brain-volume-matched template that is initialized by the rigid alignment of the respective principal brain axes (PBA). The three stages of automatic localization, segmentation and reconstruction are detailed in Sections 2.2.1 to 2.2.3, respectively⁷. All implementations are available as open-source packages⁸.

2.2.1. Localization based on Coarse Segmentation

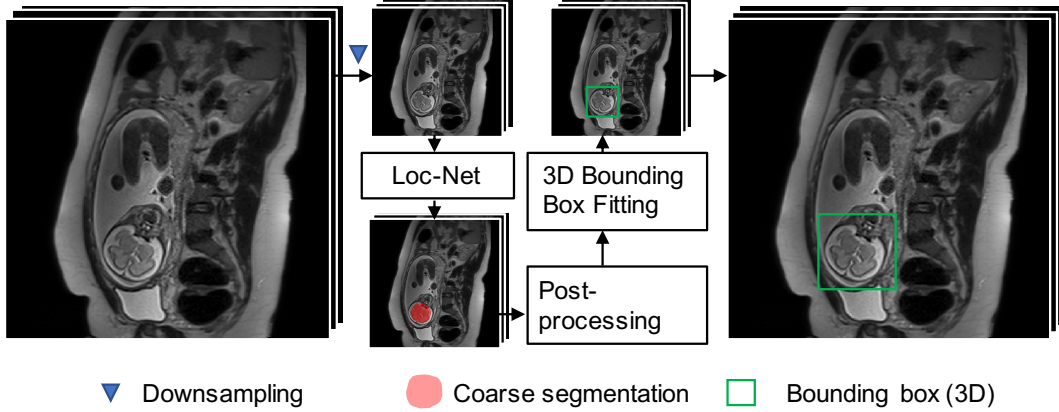


Figure 2.4.: The proposed fetal brain localization method using a CNN (Loc-Net) to obtain a coarse segmentation followed by 3D bounding box fitting.

⁷Sections 2.2.1 and 2.2.2 describe the methods for localization and segmentation of fetal brains which were developed by Guotai Wang.

⁸The automatic localization and segmentation framework FETAL_BRAIN_SEG is available at https://github.com/gift-surg/fetal_brain_seg and is integrated in the outlier-robust SRR framework NIFTYMIC (<https://github.com/gift-surg/NiftyMIC>).

Differently from traditional top-down object localization methods using sliding window classification [Criminisi et al., 2009] or bounding box regression [Gauriau et al., 2015, He et al., 2017], we use a bottom-up strategy for fetal brain localization based on a coarse segmentation by a CNN with pixel-level prediction. The advantage of such a localization method is that it gives an explainable support for the localization result and is well-suited for single-object localization. To reduce computational requirements, we apply the CNN at a down-sampled version of an input LR stack, as shown in Figure 2.4. We refer to this network for the fetal brain localization task as **Loc-Net**.

The proposed framework is theoretically amenable to different CNN models. However, as the input LR stack has a large inter-slice spacing and is potentially corrupted by motion between neighboring slices, it is more compelling to use a 2D CNN than a 3D CNN. We choose the 2D P-Net [Wang et al., 2018] for its compactness and efficiency. It consists of six blocks of convolution layers with dilated convolution [Yu and Koltun, 2016] to preserve resolution for dense prediction. The first five blocks have 2, 2, 3, 3 and 3 convolution layers respectively and they have dilation parameters of 1, 2, 4, 8 and 16, respectively. The convolution channel number for these layers is 64. Multi-scale features from these five blocks are concatenated and fed into the 6-th block which is a voxel-wise classifier with 1x1 convolution. A softmax layer is used to obtain probability-like outputs.

Let I denote a stack of slices and I_i represent the i -th 2D slice of I . To reduce the inference time and memory consumption for the localization task, we down-sample I_i to a given size, i.e., 96×96 , obtaining I'_i . We keep the through-plane resolution the same as the input stack, and use I' to denote the stack of down-sampled 2D slices. As shown in Figure 2.4, to get a 3D bounding box of the fetal brain in a stack, we first use the Loc-Net to obtain a segmentation of I' by stacking the 2D segmentations, i.e., a coarse segmentation.

With the coarse segmentation of the down-sampled stack I' , we employ two post-processing steps to reduce segmentation noise and obtain a smoother result. First, we use a 3D morphological closing and opening operation on the result of Loc-Net.

Then we select the largest connected 3D component as the post-processed coarse segmentation of the fetal brain and fit a 3D bounding box to the component as the localization result in I' . The final localization result for input I is obtained by rescaling the bounding box to the original space of I and expanding by a margin of 5 mm.

2.2.2. Fine Segmentation

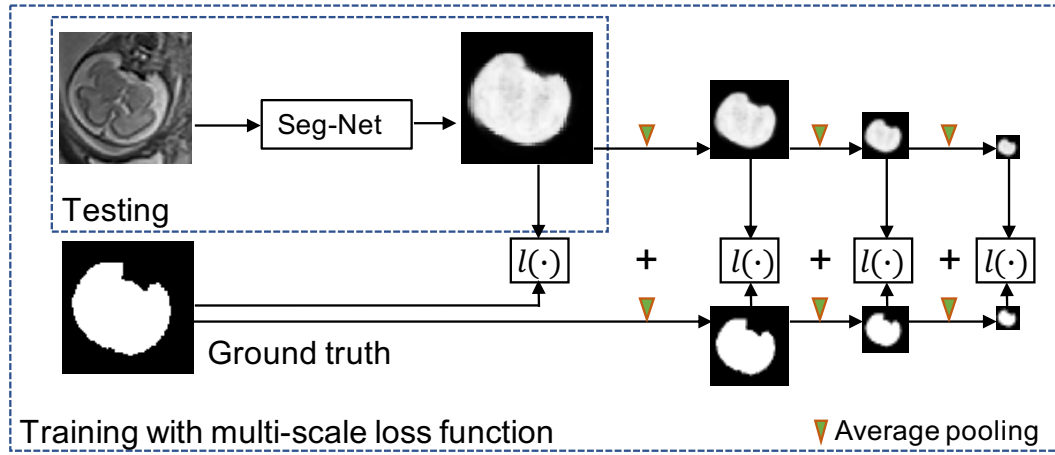


Figure 2.5.: The proposed fetal brain segmentation method using a CNN (Seg-Net) that works on the localization result. We propose to use a multi-scale loss function to train Seg-Net.

After the localization, we further obtain a fine segmentation of the fetal brain from I with a second CNN that is referred to as **Seg-Net**. It works on the ROI of the localization result of I to reduce false positives of the dense prediction. Similar to the localization step, we use the 2D P-Net structure [Wang et al., 2018] for the fine segmentation rather than a 3D network considering the inter-slice spacing and motion.

Due to the change in appearance of the fetal brain at different gestational ages and as a consequence of the presence of pathologies such as spina bifida, it is challenging to achieve robust segmentation results. We propose a multi-scale loss function for training to improve the performance of fine segmentation. The commonly adopted logistic loss and Dice loss functions for image segmentation use a sum of pixel-wise

losses [Sudre et al., 2017] and only penalize prediction errors at the finest scale, without considering the relationship between neighboring pixels at a larger scale. This potentially leads to noisy and spatially inconsistent segmentations. In contrast, dealing with the image in the scale-space representation helps to achieve more robust results, as shown by previous works inspired by the scale-space theory [Lindeberg, 1994, Hu et al., 2018].

We propose a training loss function across multiple scales as depicted in Figure 2.5. Let Y represent the pixel-wise probability prediction of an image given by a segmentation CNN and G denote the corresponding pixel-wise probabilistic ground truth. The loss function $l(Y, G)$ measures the similarity between Y and G and guides the network to obtain a segmentation as close as possible to the ground truth. It is commonly defined as a pixel-wise function for segmentation tasks. For example, the Dice loss function is defined as [Milletari et al., 2016, Sudre et al., 2017]:

$$l_{Dice}(Y, G) = 1 - \frac{2 \sum_i^N y_i g_i}{\sum_i^N y_i^2 + \sum_i^N g_i^2} \quad (2.1)$$

where N is the number of pixels. y_i represents the probability of pixel i being the foreground given by Y and g_i represents that probability given by G . Let s be a scale index ($s \in \{1, 2, \dots\}$), and Y_s and G_s be the downsampled versions of Y and G at scale s , respectively. Then the proposed multi-scale loss function is

$$L(Y, G) = \frac{1}{S} \sum_{s=1}^S l(Y_s, G_s), \quad (2.2)$$

where S is the total number of scales. Thus, the loss function $L(Y, G)$ is the average of $l(Y_s, G_s)$ across multiple scales. When $s = 1$, Y_s is the same as Y , and when $s > 1$, Y_s is a downsampled version of Y . Y_s can be obtained by down-sampling Y or applying max-pooling on Y . However, both methods cause the obtained Y_s to contain little contextual information. In contrast, Gaussian smoothing and average-pooling summarize the prediction of a local patch for more contextual information. Since average-pooling is more efficient and more straightforward to implement than Gaussian smoothing, we use average-pooling for the downscaling. Let $Pl_{avg}(\cdot)$ denote

the average-pooling operation. We set the pooling kernel size as 2×2 with stride size 2×2 . Therefore, $Pl_{avg}(\cdot)$ averages every neighboring 2×2 pixels. $Pl_{avg}(\cdot)$ is used recursively to down-scale Y and G for larger scales s :

$$Y_s = \begin{cases} Y & \text{for } s = 1 \\ Pl_{avg}(Y_{s-1}) & \text{for } s > 1 \end{cases} \quad (2.3)$$

With a larger s , Y_s and G_s encode the prediction and the ground truth at a higher level with more non-local information. Therefore, $L(Y, G)$ not only penalizes the pixel-wise difference between Y and G , but also encourages their similarity at multiple non-local scales. In this chapter, we use the Dice loss function as the loss function at each scale due to its good performance in dealing with imbalanced classes, i.e., $l(Y_s, G_s) = l_{Dice}(Y_s, G_s)$. We set the total number of scales S as 4, as shown in Figure 2.5.

2.2.3. Robust HR Volume Reconstruction Framework

The steps of the HR volume reconstruction stage are shown in Figure 2.6. We briefly list them here and further detail the main contributions in the subsequent sections. For a set of LR stacks of 2D slices acquired in multiple views, we preprocess the images using the bias field correction method [Tustison et al., 2010] as implemented in the Insight Toolkit (ITK)⁹. Using a volume-to-volume registration based on symmetric block-matching [Modat et al., 2014], all stacks are rigidly aligned with an automatically chosen target stack (more details in Section 2.3.2). Based on the brain segmentation of the target stack, all remaining, volumetrically-aligned, stacks are intensity corrected using a linear regression. An initial HR volume is obtained by applying a scattered data approximation (SDA) scheme on the LR stacks that uses an efficient discrete implementation of Nadaraya-Watson kernel regression [Vercauteren et al., 2006, Ebner et al., 2017a]. It is based on nearest neighbor sampling onto a regular grid followed by a subsequent Gaussian blurring operation for each single slice. Similarly, SDA is used to obtain a brain mask HR volume from the individual

⁹<https://itk.org>

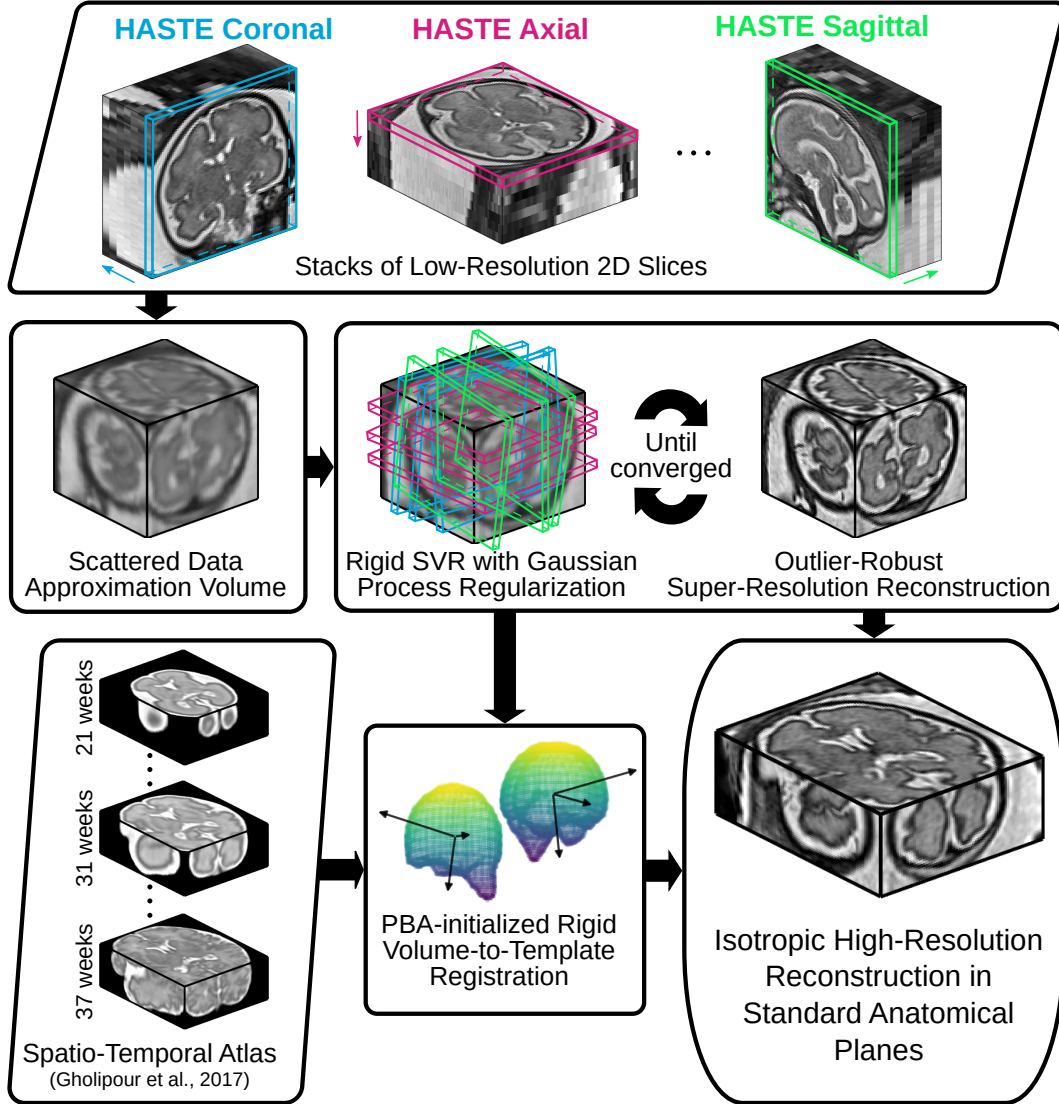


Figure 2.6.: The proposed outlier-robust HR volume reconstruction method for fetal brain MRI. As part of a two-step motion-correction/volumetric reconstruction cycle, we propose an effective robust SRR method for complete outlier rejection that relies on a single hyperparameter only and retains a linear least-squares formulation. A fast template-space alignment, which is robust also for pathological brains, is achieved by using a principal brain axes (PBA)-initialized rigid volume-to-template registration based on symmetric block-matching.

LR stack masks. Then, an updated HR volume is obtained through a two-step iterative registration-reconstruction approach [Rousseau et al., 2006, Gholipour et al., 2010b]. In each iteration, the rigid registration step registers the slices to the HR volume constructed in the previous iteration for motion correction constrained by the

respective slice and HR brain mask. Using Gaussian Process Regularization based on a t -distribution noise model between each sub-stack of an interleaved acquisition, we investigate a robust statistical model as a temporal prior for the slice transformations. Subsequently, the reconstruction step constructs an HR volume and brain mask from the aligned slices and segmentations, respectively. After reconstruction in the subject's space, we rigidly align the HR volume to a spatiotemporal atlas of normal brains [Gholipour et al., 2017] to obtain the reconstruction in the standard anatomical planes.

Robust Motion Correction using Gaussian Process Regularization

To obtain more robust registration outcomes, we want to reduce the likelihood of occurring misregistrations for individual slices. Each slice of a sub-stack in fetal MRI is acquired with fast temporal resolution and therefore leads to a high sampling of the continuous underlying physical motion of the fetus. Thus, we hypothesize that slices within each sub-stack of an interleaved acquisition are likely to have similar transformation parameters. Let $(\theta_{k,1}, \dots, \theta_{k,6}) \in \mathbb{R}^6$, $k = 1, \dots, K$, describe the parameters associated with the 3D rigid registration transformations of the K slices for a given sub-stack of the M -interleaved acquisitions. We assume that the r -th transformation parameter changes smoothly for the slices across a given sub-stack, i.e. we assume $\theta_{k,r}$ changes smoothly for $k = 1, \dots, K$ for each fixed r . We interpret the estimated registration parameters $(\theta_{k,r})_{k=1}^K$ as the sampled observations from a Gaussian process [MacKay, 1998] that are temporally dependent on a Brownian motion. For increased robustness to outliers, we assume a t -process prior $\mathcal{TP}(\nu, \tilde{\mu}, \tilde{\sigma})$ with mean $\tilde{\mu}$ and standard deviation $\tilde{\sigma}$ as the noise model as a generalization of a Gaussian process prior $\mathcal{GP}(\tilde{\mu}, \tilde{\sigma})$, whereby ν controls how heavy-tailed the process is. Thus, we can write

$$\begin{aligned} z_k &\sim \mathcal{N}(z_{k-1} + \mu, (1 - \lambda)\sigma^2) \quad \text{for } k = 2, \dots, K \\ \theta_{k,r} &\sim \mathcal{TP}(\nu, z_k, \lambda\sigma^2) \end{aligned} \tag{2.4}$$

whereby z_k is the hidden Brownian motion, i.e. the robust registration parameter value estimate associated with the observed, noisy value $\theta_{k,r}$, μ the expectation of the random offset between z_k and z_{k-1} , and σ^2 the total variance. Thus, (2.4) is split into two parts where the variance of the hidden Brownian motion is $(1 - \lambda)\sigma^2$ and the variance of the observation noise is $\lambda\sigma^2$ whereby $\lambda \in (0, 1)$ balances the respective contributions. The higher λ , the more variance is assigned to the noise which therefore leads to a higher smoothness in $(z_k)_{k=1}^K$. To reduce the influence of noise, a maximum a-posteriori estimator is used to obtain the values of the hidden process z_k as a regularized version of $\theta_{k,r}$.

Outlier-Robust Super-Resolution Reconstruction

After each SVR step of the two-step registration-reconstruction iteration i , an SRR step is used to recover the most likely HR volume \mathbf{x}^i that satisfies the slice acquisition model (1.1) [Gholipour et al., 2010b]

$$\mathbf{y}_k^i = \mathbf{A}_k^i \mathbf{x}^i + \mathbf{n}_k^i \quad (2.5)$$

where \mathbf{y}_k^i is the k -th slice in a stack. \mathbf{A}_k^i represents the image acquisition process including rigid transformation, slice selection, blurring according to the PSF, and down-sampling. \mathbf{n}_k^i denotes the vector of observed noise. The intensity of each voxel in an LR slice is therefore influenced by a certain neighborhood of this voxel within an HR volume \mathbf{x} given by the assumed PSF that is specific to the slice profile of the MR acquisition [Liang and Lauterbur, 2000]. For SST2W sequences, a common approximation is given by a slice-aligned 3D Gaussian function that depends on the in- and through-plane resolution of the LR slice [Jiang et al., 2007, Kuklisova-Murgasova et al., 2012]. The position and orientation of the slice (and PSF) within the volume is estimated in the rigid SVR step. In order to prevent misregistered or artefact-corrupted outlier slices from affecting the reconstructed HR volume, we propose a

robust SRR with outlier rejection in a maximum a-posteriori (MAP) formulation

$$\mathbf{x}^i := \arg \min_{\mathbf{x} \geq 0} \left(\sum_{k \in \mathcal{K}_\beta^i} \frac{1}{2} \|\mathbf{y}_k^i - \mathbf{A}_k^i \mathbf{x}\|_{\ell^2}^2 + \frac{\alpha}{2} \|\nabla \mathbf{x}\|_{\ell^2}^2 \right) \quad (2.6)$$

where $\alpha \geq 0$ denotes a regularization parameter, ∇ the differential operator, and \mathcal{K}_β^i a set of indices of inlier slices

$$\mathcal{K}_\beta^i := \{1 \leq k \leq K : \text{Sim}(\mathbf{y}_k^i, \mathbf{A}_k^i \mathbf{x}^{i-1}) \geq \beta\} \quad (2.7)$$

that demonstrate high agreement with their simulated counterparts projected from the previous SRR iterate using a similarity measure Sim and a threshold $\beta > 0$. Thus, slices with a value of $\text{Sim}(\cdot)$ lower than β are regarded as outliers and rejected in the SRR step. More complex SRR models have been proposed in addition to the MAP formulation including modifying (2.6) to rely on robust M-estimator [Gholipour et al., 2010b] and total variation formulations [Toubier et al., 2015]. However, while they substantially increase the computational cost, in our experience, they tend to show little improvement in the obtained reconstruction quality in case of appropriate motion correction¹⁰. Assuming a fixed \mathcal{K}_β^i , we obtain a convex SRR problem with complete outlier rejection in a linear least-squares formulation which we solve using *matrix-free* operations [Diamond and Boyd, 2015, Ebner et al., 2017a]. We use a dedicated linear least-squares solver to deal with this large linear system whereby positivity is enforced by clipping negative values.

Furthermore, we create an HR brain mask by applying the fast SDA approach on the motion-corrected inlier slice masks which is used for both motion correction and the labelling of inlier-slices in (2.7) in the subsequent iteration.

Reconstruction in Standard Anatomical Planes

Obtaining the HR fetal brain reconstructions in the standard anatomical planes can facilitate brain studies and is typically favored for the clinical assessment by clini-

¹⁰Associated experiments for MRCP SRR are summarized in Section 3.4.2

Table 2.1.: Information of the datasets used for the experiments. The number of available subjects and stacks are listed for training, validation and testing, respectively. Gestational age (GA) is stated as mean and standard deviation.

	Group A	Group B1	Group B2
Pathology	Normal	Spina bifida (pre-surgical)	Spina bifida (post-surgical)
GA (weeks)	29.51±4.46	23.47±2.86	25.73±1.28
Subjects	(26, 4, 7)	(12, 4, 16)	(0, 0, 16)
Stacks	(78, 12, 44)	(36, 12, 119)	(0, 0, 105)

cians. To define the template space, we deployed the spatiotemporal atlas¹¹ [Gholipour et al., 2017] which was constructed from 81 normal fetuses scanned between 19 and 39 weeks of gestation. Rigid registration can be used to align the subject-space SRR to a template. However, given the substantial morphological differences in brain volume and shape between pathological and normal fetuses, a direct registration approach is likely to get stuck in local minima leading to an incorrect template space alignment. To avoid this problem, we propose to use an initialized transformation that is based on the rigid alignment of fetal brain masks only. Using principal component analysis, we first rigidly align the principal brain axes (PBA) of the template and HR brain masks whereby the template is selected based on brain-volume matching. Following the PBA-based initialization, we perform a 3D rigid registration based on block-matching [Modat et al., 2014]. For increased robustness, all four permutations of the right-handed bases of principal eigenvectors are tested and the best registration transform is selected as determined by the normalized mutual information similarity between warped SRR and template.

After the 3D rigid registration, we use an additional SRR step to obtain the HR volume in the template space, considering that the resampling process during the 3D rigid registration may affect the image quality.

2.3. Experiments and Results

2.3.1. Data

The automatic reconstruction framework was applied to routinely acquired clinical images of fetuses with normal brains, yet scanned for other anomalies, and fetuses with spina bifida (SB) that were scanned at University Hospitals KU Leuven as reported in [Aertsen et al., 2019]. Access to anonymized images was facilitated through the GIFT-Cloud platform for data sharing [Doel et al., 2017]. For normal fetuses, 134 stacks from 37 individuals were scanned at the gestational age (GA) of 29.51 ± 4.46 weeks (“normal” group A). 32 fetuses with spina bifida were scanned before fetal surgical closure at a GA of 23.47 ± 2.86 weeks (“pre-surgery” group B1), and 16 of them were additionally scanned after fetal surgical closure at a GA of 25.73 ± 1.28 weeks (“post-surgery” group B2). Details of the dataset are summarized in Table 2.1. For each study, 3 to 9 SST2W stacks in at least three different orientations were collected with pixel size 0.39 mm to 1.48 mm and slice thickness 2.50 mm to 4.40 mm. All images were acquired with no slice overlap nor gap. For the purpose of testing the robustness of our proposed framework, we also kept heavily motion- and artefact-corrupted stacks and also images where brains were only partially scanned.

For the fetal brain detection and segmentation set-up, 78 stacks of 26 patients from Group A and 36 stacks of 12 patients from Group B1 were used for training, and 12 stacks of 4 patients from Group A and 12 stacks of 4 patients from Group B1 were used for validation. The remaining images were used for testing, as shown in Table 2.1. Manual segmentations of the fetal brains on the 2D slices were used as the ground truth for the segmentation task, and the bounding box of the manual segmentation was extended by 5 mm to be used as the ground truth for fetal brain localization. We normalized the intensity of each stack by its mean and standard deviation.

¹¹http://crl.med.harvard.edu/research/fetal_brain_atlas

2.3.2. Implementation Details

Our CNNs were implemented in TensorFlow¹² using NiftyNet¹³ [Li et al., 2017a, Gibson et al., 2018]. For the training of Seg-Net, we set the number of scales S to 4 in (2.2) and employed Dice loss as the loss function used in each individual scale. The detection/segmentation experiments were implemented with an NVIDIA GeForce GTX 980 GPU. For both Loc-Net and Seg-Net, we used Adaptive Moment Estimation (Adam) [Kingma and Ba, 2015] for training, with initial learning rate 10^{-3} , batch size 10, weight decay 10^{-7} and 10^4 iterations.

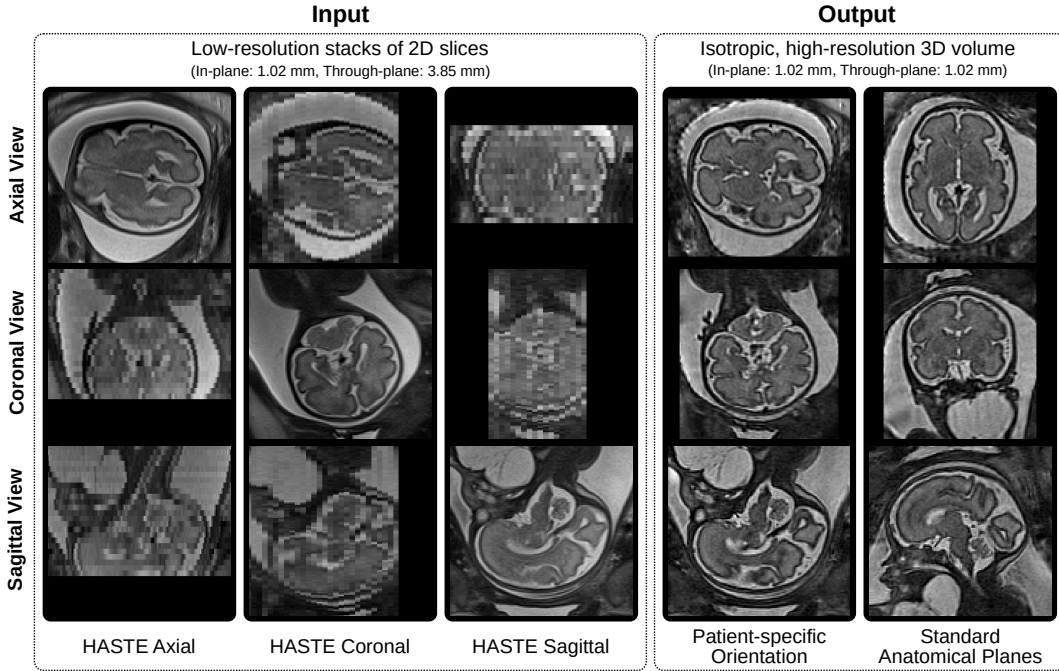


Figure 2.7.: Volumetric reconstructions in patient-specific orientation and standard anatomical planes. The image is adapted from the GitHub-page of NIFTYMIC and shows the HR brain visualizations of a fetal neck mass subject [Ebner et al., 2016] using a 1 mm isotropic reconstruction grid (5 input LR stacks, GA = 30 weeks).

The HR volume reconstruction algorithm was applied to the testing data of Group A (normal), B1 (pre-surgical spina bifida) and B2 (post-surgical spina bifida) cases using the obtained automatic segmentation results to guide the rigid SVR step. The reconstruction was performed with three cycles of the two-step motion correction

¹²<https://www.tensorflow.org>

¹³<http://niftynet.io>

and robust volumetric reconstruction iterations. For automatic target stack selection, we empirically chose the stack with (estimated) brain volume closest to 70% of the median brain volume using the automatically segmented (Seg-Net) brain masks. This approach was chosen for simplicity to define a target stack that shows good brain coverage but avoids stacks with unrealistically high brain volume estimates due to false-positive segmentations or heavy motion-corruption. However, and as shown in the results section, this may still lead to a target stack with substantial motion artefacts suitable to test the robustness of our algorithm also for challenging cases. The orientation of the subject space is defined by the target stack whereby the reconstruction grid is obtained by extending the bounding box of the union of the volumetrically aligned stack masks by 10 mm in each direction. For the respective volume-to-volume (and volume-to-template) rigid registrations we deployed the symmetric block-matching algorithm REGALADIN as part of NIFTYREG¹⁴ [Modat et al., 2014]. For the rigid slice-to-volume motion correction steps we used ITK with normalized cross-correlation (NCC) to guide the registrations, whereby a conjugate gradient descent solver in combination with line-search for the learning rate was used as the optimizer. The implementation of the SDA approach was based on the Young & Van Vliet recursive Gaussian smoothing filter¹⁵[Vidal-Migallon et al., 2013]. We empirically chose a standard deviation of 1 for both the HR volume and brain mask HR volume SDAs. For robust SVR, we used GPR to obtain regularized transformation parameters within each sub-stack of the interleaved acquisition with a default smoothing parameter $\lambda = 0.5$ in (2.4). We set the expectation of the random offset μ to zero to avoid any systematic bias in the parameter estimates. The remaining parameters in (2.4) were drawn from distributions with manually set parameters. The total variance σ^2 was sampled from an exponential distribution with parameter equal to 1, i.e $\sigma \sim \exp(1)$. The degrees of freedom ν for the \mathcal{TP} prior were drawn from a Gamma-distribution with shape $\alpha = 2$ and rate $\beta = 1$, i.e. $\nu \sim \Gamma(2, 1)$. The maximum likelihood estimation of z_k in (2.4) was implemented

¹⁴<https://github.com/KCL-BMEIS/niftyreg>

¹⁵<https://github.com/Inria-Asclepios/SmoothingRecursiveYvGaussianFilter>

using the probabilistic Python package PyMC3¹⁶ [Salvatier et al., 2016]. After performed GPR step, we performed another SVR step to refine the respective rigid slice transformations as initialized by the GPR estimate. GPR was applied only for the first SVR cycle; all subsequent motion-correction steps were performed using rigid slice-to-volume registration only. To parametrize the outlier-robust SRR approach, experiments were performed to investigate its sensitivity to the outlier-threshold β and the input fetal brain masks. Associated comparisons are summarized in Section 2.3.5. By choosing NCC as the similarity measurement function $\text{Sim}(\cdot)$ in (2.7), a good balance between conservative slice retention and effective outlier rejection was found for $\beta = 0.8$. For the experiments, we set the threshold value β to be 0.5, 0.65 and 0.8 per iteration to account for increasing accuracy in (2.6), respectively. The matrix-free implementation of the forward operator \mathbf{A}_k^i in (2.6) (and its adjoint) was done by extending the resampling operator in ITK to allow for oriented Gaussian filtering¹⁷ representing the oriented PSF kernel whereby the SCIPY LSMR solver was used for the numerical optimization in (2.6). The isotropic resolution of the HR volume was set to match the final template space resolution of 0.8 mm [Gholipour et al., 2017] for both subject and template space reconstructions. The regularization parameter $\alpha = 0.01$ was determined by visual assessments supported by L-curve studies [Hansen, 2001]. We reconstructed the entire field of view for both subject and template spaces from the brain-motion corrected slices to provide anatomical context beyond the brain¹⁸. Our Python code, including both the automatic brain segmentation tool FETAL_BRAIN_SEG¹⁹ and the outlier-robust SRR framework NIFTYMIC²⁰, is publicly available. An overview of the reconstructions obtained by NIFTYMIC in the patient-specific orientation and standard anatomical planes for a fetal neck mass subject [Ebner et al., 2016] is provided in Figure 2.7.

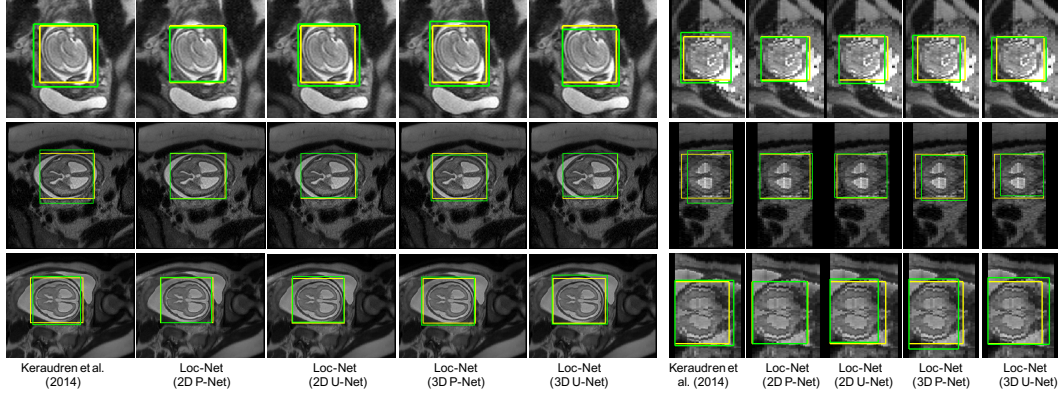


Figure 2.8.: Visual comparison of different methods for fetal brain localization. The three rows show examples from Group A (controls), B1 (pre-surgical spina bifida), and B2 (post-surgical spina bifida), respectively. Column 1-5: in-plane. Column 6-10: through-plane. Yellow: ground truth. Green: detection results.

Table 2.2.: Quantitative evaluation of different methods for fetal brain localization. The best values are shown with bold font. * denotes significant difference from the other values based on a paired t -test (p -value < 0.05).

	IoU (%)			Centroid distance (mm)			Stack-level runtime(s)
	Group A	Group B1	Group B2	Group A	Group B1	Group B2	
[Keraudren et al., 2014]	72.38 \pm 12.88	72.87 \pm 10.37	69.18 \pm 11.73	6.70 \pm 4.04	7.31 \pm 4.58	7.68 \pm 2.83	15.03 \pm 3.54
Loc-Net (2D P-Net)	86.54\pm5.87*	84.74\pm5.55*	83.67\pm5.04*	3.51\pm1.85*	3.70\pm2.10*	4.51\pm2.45*	2.35 \pm 1.02
Loc-Net (2D U-Net)	83.22 \pm 6.94	81.26 \pm 6.80	78.32 \pm 8.81	4.25 \pm 2.33	4.46 \pm 1.98	5.38 \pm 2.59	2.12\pm1.03
Loc-Net (3D P-Net)	82.68 \pm 8.14	78.26 \pm 8.87	76.58 \pm 12.74	4.46 \pm 2.29	4.77 \pm 2.76	5.54 \pm 2.83	2.49 \pm 1.05
Loc-Net (3D U-Net)	80.02 \pm 10.57	76.67 \pm 9.83	75.88 \pm 12.03	5.73 \pm 3.08	5.89 \pm 3.06	5.76 \pm 3.64	2.31 \pm 1.08

2.3.3. Localization Results

For the choice of network structure of Loc-Net, we compared 2D P-Net with 2D U-Net, and also compared them with 3D P-Net [Wang et al., 2018] and 3D U-Net [Çiçek et al., 2016] to investigate whether 2D or 3D networks are more suitable for uncorrected fetal MR image stacks. These networks were all implemented using NiftyNet. In addition, we compared Loc-Net using these network structures with the method described in [Keraudren et al., 2014]²¹ that is based on classification of image regions using SIFT features and combined with prior knowledge of brain size

¹⁶<https://docs.pymc.io/intro.html>

¹⁷https://github.com/gift-surg/ITK_NiftyMIC

¹⁸As we only correct for rigid brain motion, the surrounding maternal/fetal tissue reconstruction may be of limited anatomical accuracy.

¹⁹https://github.com/gift-surg/fetal_brain_seg

²⁰<https://github.com/gift-surg/NiftyMIC>

²¹<https://github.com/BioMedIA/IRTK>

and shape based on gestational age.

Figure 2.8 shows the fetal brain localization results for three cases from Group A, B1 and B2 respectively. In the first case, the centroid of the bounding box obtained by [Keraudren et al., 2014] is close to that of the ground truth. However, the size of the localization result is larger than that of the ground truth. In contrast, the result of our Loc-Net with 2D P-Net matches better with the ground truth. It can also be observed that the result of 2D P-Net is better than that of the other three networks. In the second case, the in-plane visualization shows that similar results are achieved by the 2D and 3D networks. However, the through-plane visualization shows that 2D P-Net and 2D U-Net outperform their 3D counterparts.

Quantitative evaluations of these localization methods are shown in Table 2.2. We calculated the Intersection over Union (IoU) score and centroid distance between the localized 3D bounding box and the localization ground truth. Our Loc-Net with 2D P-Net achieved average IoUs of 86.54%, 84.74% and 83.67% for these three groups, respectively, and they are more than 10% higher than that of [Keraudren et al., 2014]. The runtime measurements show that Loc-Net is more than 6 times faster than [Keraudren et al., 2014]. Table 2.2 also shows that 2D P-Net outperforms 2D U-Net when used as Loc-Net, and both of them achieve better localization accuracy than their 3D counterparts.

2.3.4. Segmentation Results

Table 2.3.: Quantitative evaluation of fetal brain segmentation. The best values are shown with bold font. A * denotes significant improvement compared with the corresponding baseline without ML based on a paired t -test (p -value < 0.05). The runtime of Seg-Net is a sum of time consumed by localization and segmentation.

	Dice (%)			Hausdorff (mm)			Stack-level runtime(s)
	Group A	Group B1	Group B2	Group A	Group B1	Group B2	
[Salehi et al., 2018]	90.48±4.61	90.12±4.19	88.57±4.24	12.01±6.87	14.75±7.51	12.08±6.63	1.98±0.76
Seg-Net (2D P-Net)	92.27±3.83	91.56±3.33	90.93±4.95	11.26±8.25	10.93±5.66	9.83±5.29	3.65±1.34
Seg-Net (2D P-Net) + ML	93.21±4.02*	93.87±2.79*	92.94±4.14*	9.67±6.55*	6.94±4.29*	7.84±3.26*	3.66±1.33
Seg-Net (2D U-Net)	92.50±2.85	91.10±3.41	91.48±4.31	10.41±6.34	10.91±6.80	9.76±6.21	3.40±1.35
Seg-Net (2D U-Net) + ML	93.45±2.99*	91.57±2.24	92.49±3.61*	9.13±5.28	7.27±3.89*	7.71±4.26*	3.40±1.35

We compared using 2D P-Net and 2D U-Net as the Seg-Net. As a baseline, both of them were trained with the Dice loss with a scale $s = 1$, i.e. at the input resolution.

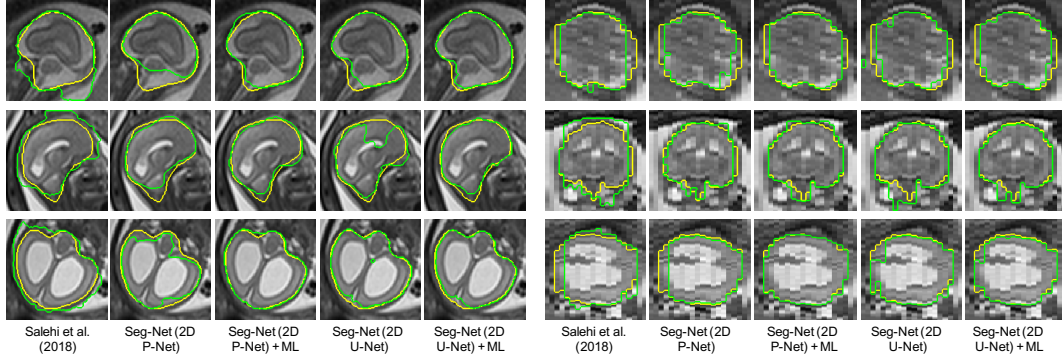


Figure 2.9.: Visual comparison of different methods for fetal brain segmentation. The three rows show examples from Group A (controls), B1 (pre-surgical spina bifida), and B2 (post-surgical spina bifida), respectively. Column 1-5: in-plane. Column 6-10: through-plane. Yellow: ground truth. Green: segmentation results.

Then we trained these networks with the proposed multi-scale (ML) Dice loss. These four variants are referred to as Seg-Net (2D P-Net), Seg-Net (2D U-Net), Seg-Net (2D P-Net) + ML, Seg-Net (2D U-Net) + ML respectively. All of them take the same localization result of Loc-Net (2D P-Net) as input for a fair comparison. In addition, we compared them with the method described in [Salehi et al., 2018] that applies 2D U-Net to the whole input image for segmentation without a localization stage. We followed their implementation available online²² and trained the model from scratch with our training images.

Figure 2.9 presents a visual comparison of these segmentation methods for images from Group A, B1 and B2 respectively. It shows that the method of [Salehi et al., 2018] tends to generate false positives in tissues surrounding the fetal brain. In contrast, the variants of Seg-Net based on localization results achieve more accurate segmentation with reduced false positives. The results of Seg-Net (2D P-Net) and Seg-Net (2D U-Net) have some under-segmentations and unsmoothed contours. By training with the proposed multi-scale loss function, their corresponding results are more spatially consistent and accurate.

Table 2.3 presents quantitative evaluations of these segmentation methods for the fetal brain. We calculated the Dice score and Hausdorff distance between the seg-

²²<https://bitbucket.org/bchradiology/u-net/src>

mentation results and the pixel-level ground truth. It shows that all the Seg-Net variants achieved a higher segmentation accuracy than [Salehi et al., 2018]. Seg-Net (2D P-Net) + ML achieved average Dice scores of 93.21%, 93.87% and 92.94% for Group A, B1 and B2 respectively, and it significantly outperformed Seg-Net (2D P-Net) that was trained without ML. The total average time of our CNN-based localization and segmentation steps was less than four seconds for one stack.

2.3.5. Outlier-Robust SRR Results

For the experiments, we computed the HR volume reconstructions using various methods: 1) the automatic localization results based on Loc-Net (2D P-Net), 2) the automatic fine segmentation results obtained by Seg-Net (2D P-Net), and 3) manual segmentation results. These three variants are denoted as SRR (L), SRR (S) and SRR (M), respectively. Additionally, we applied the state-of-the-art toolkit provided by [Kainz et al., 2015b]²³ using the manual segmentations as input masks, denoted as Kainz et al. (M)²⁴.

First, we performed experiments to compare the impact of using GPR in the first slice-to-volume motion-correction cycle. Figure 2.10 illustrates how GPR can detect apparent outliers after the SVR step and therefore lead to more consistent transformations of slices within a sub-stack. In Figure 2.11, a comparison on the overall number of retained, i.e. non-rejected, slices used for the HR reconstruction outcomes for both SRR (M) and SRR (S) for all 39 cases. Despite reducing the number of slice rejections, GPR leads to an only minimal improvement of achieving approximately 0.74 ± 1.29 and 0.56 ± 1.50 more slices for SRR (M) and SRR (S) for the final volumetric reconstruction step, respectively. Therefore, our experiments suggest that using GPR does not lead to a demonstrable improvement in obtaining more robust slice motion-correction transformations, unfortunately. Thus, we perform a “standard” slice-to-volume registration approach for motion correction and do not include the GPR step in the volumetric reconstruction algorithm for the remainder

²³<https://github.com/bkainz/fetalReconstruction>

²⁴For the reconstructions, we used the CPU version for both SVR and SRR as the GPU-accelerated approach tends to produce blurrier SRR outcomes in our experience. Therefore, the resulting benchmark method reduces to the approach as described in [Kuklisova-Murgasova et al., 2012].

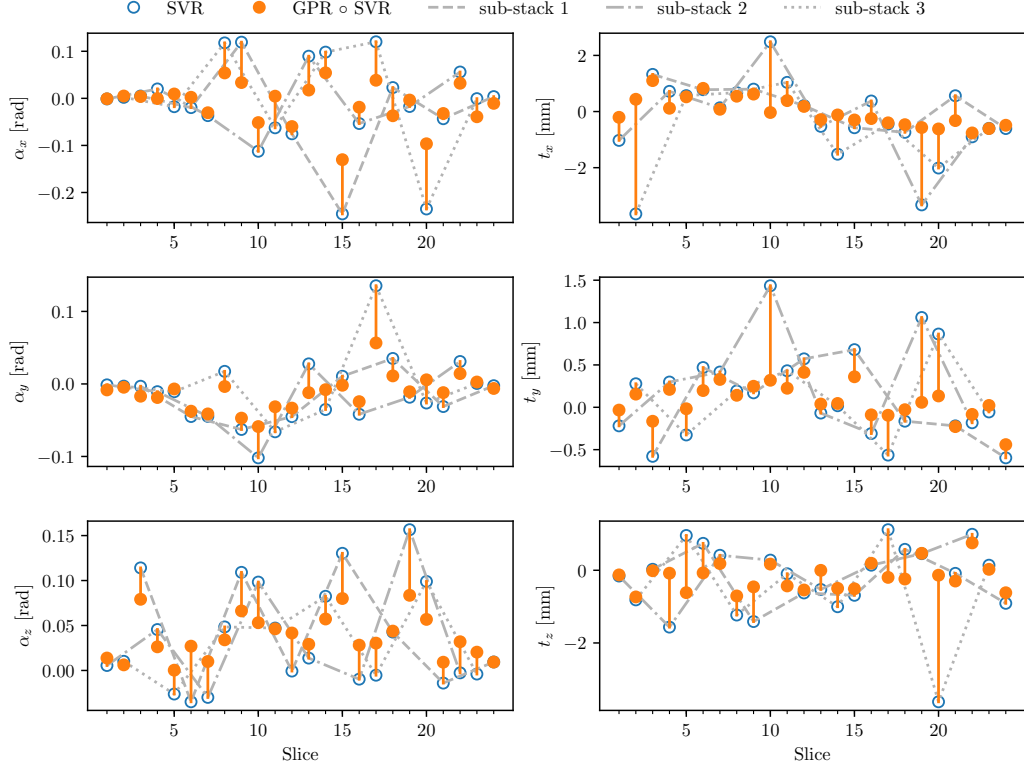


Figure 2.10.: Original SVR registration parameter values and the outcome after using GPR for an example stack that was acquired with an interleave of 3. The three rotation parameters are denoted as α_x , α_y and α_z , and the three translation parameters as t_x , t_y and t_z , respectively. Especially for the translational parameters t_x , t_y and t_z , GPR visibly leads to a more continuous change in parameter values by updating apparent outliers as compared to adjacent slices acquired within the same sub-stack.

of this chapter.

Experiments were performed to investigate the sensitivity of the proposed HR reconstruction method to the outlier-threshold β and the input fetal brain masks. Figure 2.12 shows how the outlier-threshold β impacts the number of retained slices used for solving the SRR problem (2.6). The higher β the higher and less variant the volumetric self-consistency becomes as expressed by the slice similarities $\text{Sim}(\mathbf{y}_k^i, \mathbf{A}_k^i \mathbf{x}^i)$ after the final iteration $i = 3$. However, the number of slice rejections substantially increases beyond $\beta = 0.8$ whereas the NCC of measured slice similarities plateaus. To strike a balance between conservative slice retention and effective

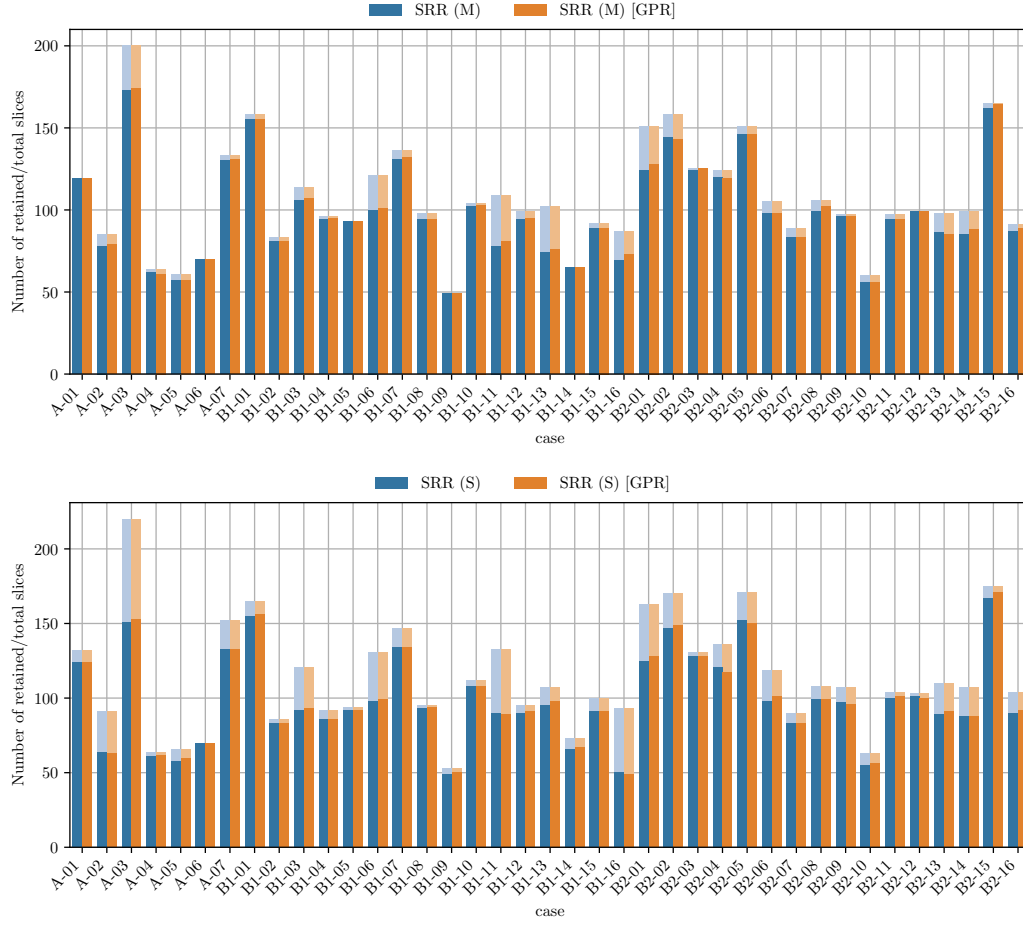


Figure 2.11.: Resulting number of retained slices used for the HR reconstruction outcomes using both manual and the automatically segmented masks, i.e. SRR (M) and SRR (S), based on either original SVR or SVR combined with GPR, respectively. Results are provided for 39 cases including 6 normal (A), 16 pre-surgical (B1) and 16 post-surgical (B2) cases. In general, GPR tends to provide only minimal improvement as the final figures of retained slices are comparable. On average, the GPR leads to an increased retention of only 0.74 ± 1.29 and 0.56 ± 1.50 slices for SRR (M) and SRR (S), respectively.

outlier rejection we select $\beta = 0.8$ as the defining outlier-threshold for our method. Figure 2.13 illustrates the slice rejection performance based on different input masks. In case of manual masks, typically well beyond 90 % of all slices are retained for the volumetric reconstruction. Automatically obtained masks lead to a higher rate of slice rejections but result in an overall comparable number of remaining slices for the SRR step. Thus, slices with false-positives segmentations are automatically detected

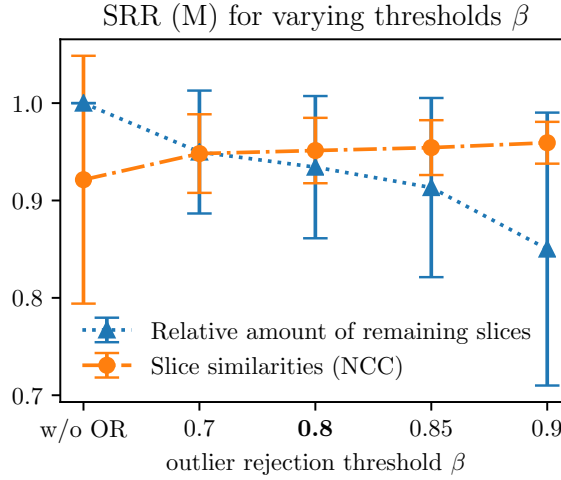


Figure 2.12.: Impact of outlier threshold on slice retention and slice similarities. Impact of outlier rejection threshold β on the slice similarities $\text{Sim}(\mathbf{y}_k^i, \mathbf{A}_k^i \mathbf{x}^i)$ for SRR (M) for the respectively remaining \mathcal{K}_β^i -slices at iteration $i = 3$. The error bars indicate the mean and standard deviation. A good balance between a high number of retained slices and high slice similarity (indicating a good self-consistency of the obtained SRR) appears to be around $\beta = 0.8$.

and rejected by the SRR algorithm.

All of the 39 cases of the groups A (7), B1 (16) and B2 (16) were used for analysis as at least one of the reconstruction methods provided a successful reconstruction in the subject-space. In Table 2.4 we demonstrate the effectiveness of the proposed template-space alignment step. For comparison purposes, we also provided the number of successful template-space alignments using NIFTYREG, FLIRT²⁵[Jenkinson et al., 2002] and their respective compositions. Overall, only two cases failed at the template-space alignment step for SRR (S) (one each for B1 and B2) for our proposed method whereas all alignments were successful for SRR (M). All seven cases of group A were successfully reconstructed and aligned. Therefore, the success rate of our framework for all the groups was 37/39 (and 39/39 when the template-space alignment is not considered). The failed template-space alignments were manually initialized for SRR (S) so that in total all 39 cases were available for the following evaluations. Some visual comparisons of the obtained SRR (S) in the template space

²⁵<https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FLIRT>

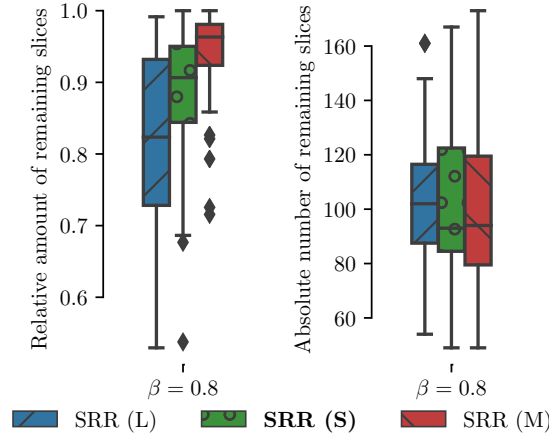


Figure 2.13.: Impact of outlier threshold for different inputs masks. Comparison of remaining slices given by \mathcal{K}_β^i at iteration $i = 3$ for the outlier-robust SRR algorithm using different input masks. Automatic segmentations lead to a higher rate of slice rejections compared to using the manual input masks but result in an overall comparable number of remaining slices. Thus, slices with false-positive segmentations are automatically detected and rejected by the SRR algorithm.

along with the HR mask reconstructions for different input masks are provided in Figure 2.14.

Figure 2.15 presents a visual comparison of the obtained SRRs for a B1 (pre-surgical spina bifida) and an A (normal) case in the subject space. Despite the challenging target stacks due to intra-stack motion, in-plane blur and intensity artefact corruption, successful reconstructions were obtained using the proposed outlier-robust SRR method. In particular, the HR visualizations for SRR (M) and SRR (S) appear visually almost indistinguishable.

For quantitative evaluations, we used $\text{Sim}(\mathbf{y}_k^i, \mathbf{A}_k^i \mathbf{x}^i)$ after the final SVR-SRR iteration ($i = 3$) to measure the similarities between the motion-corrected slices of the input LR stacks and their corresponding simulated slices from the reconstructed HR volume. We present structural similarity index measure [Wang et al., 2004] (SSIM) and peak-signal-to-noise ratio (PSNR) measurements for the comparisons here. Alternative similarity measures (NCC; normalised mutual information, NMI; root mean squared error, RMSE; mean absolute error, MAE) were also generated and are presented in Figure A.1 for the sake of manuscript conciseness. Figure 2.16

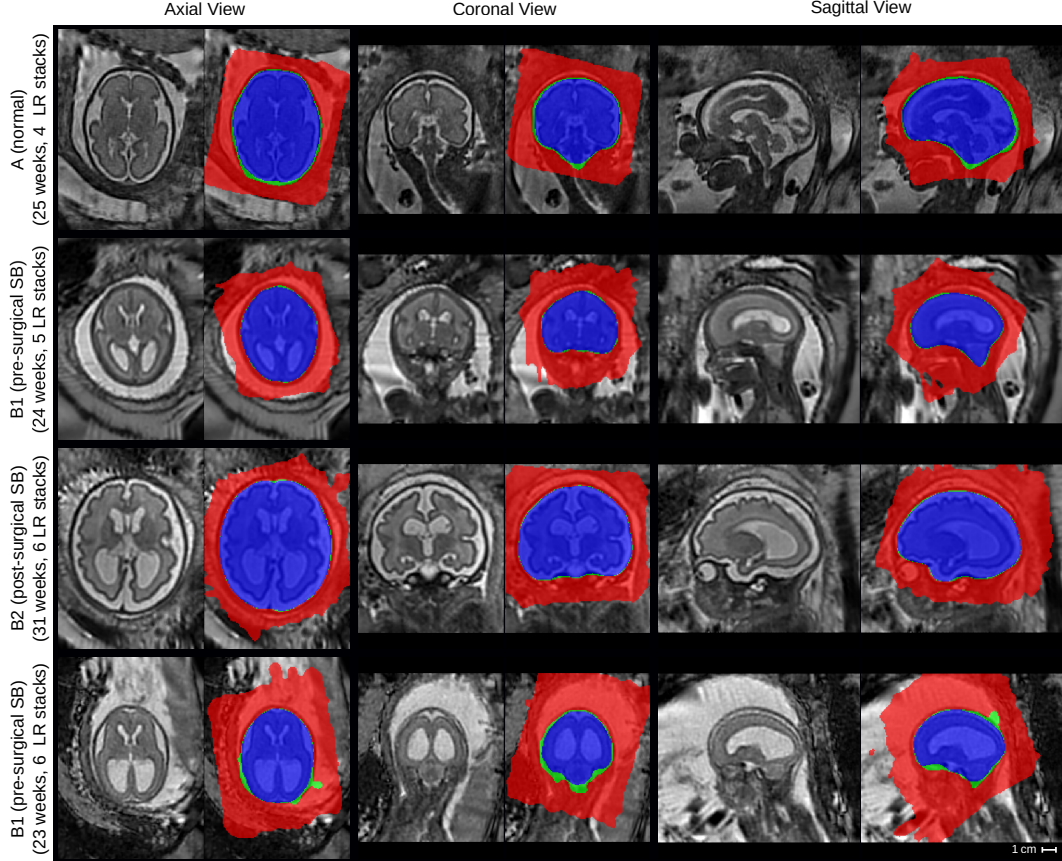


Figure 2.14.: Comparison of SRR (S) with overlaid SRR (L)/(M)/(S) HR masks obtained using either the manual masks (SRR (M); blue colour), the automatic segmentations by Seg-Net (SRR (S); differences to SRR (M) in green colour) or the localization results by Loc-Net (SRR (L); differences to SRR (M)/(S) in red colour). The respective visualizations of SRR (S) were obtained by reconstructing the entire template-space field of view using the brain-motion corrected slice transformations transformed into the template space. The last row shows the only B1-case that failed in the template-space alignment step for SRR (S), see Table 2.4; the final alignment was obtained after manual re-initialization of the volume-to-template registration.

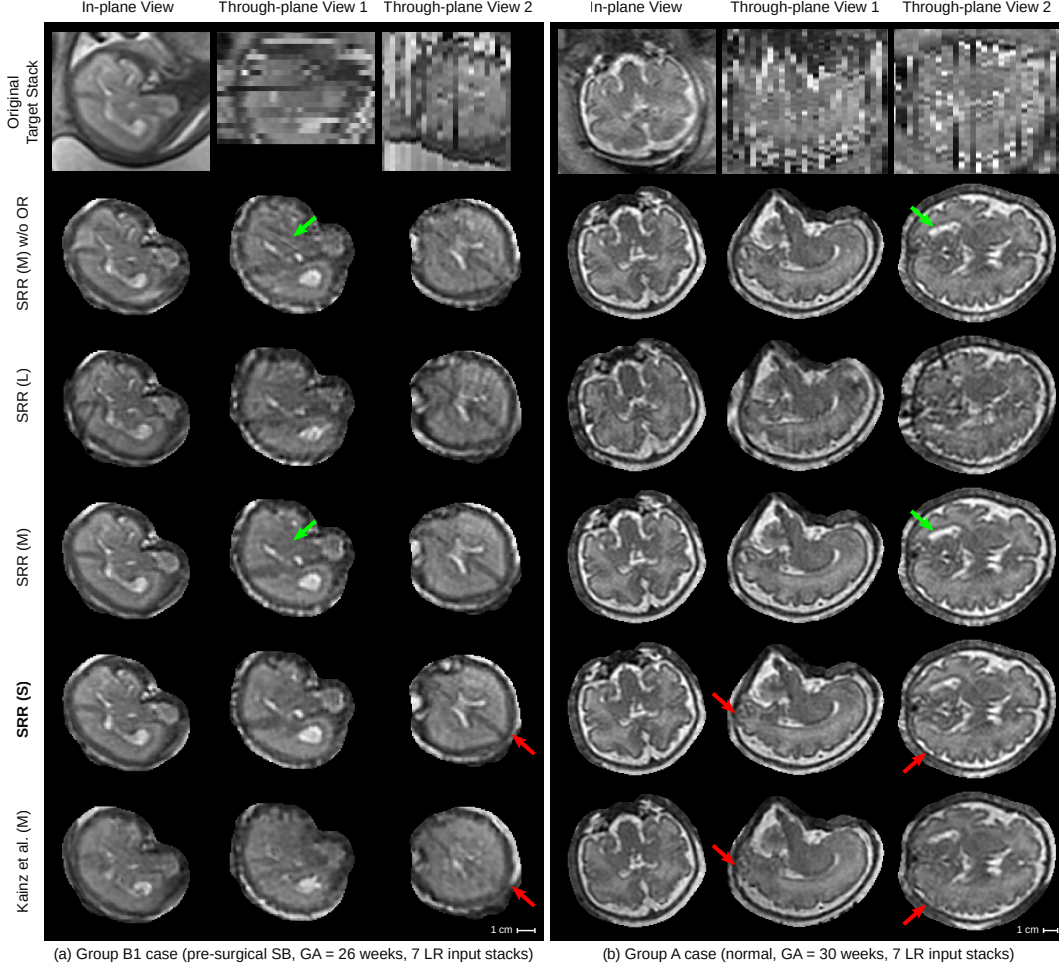


Figure 2.15.: Qualitative comparison of reconstruction methods in the subject space. Visual comparisons of different reconstruction methods for a B1 (left) and an A (right) case where challenging target stacks were (automatically) selected. An additional comparison of the group A case in the template space is shown in Figure 2.20. Dilated SRR (M) masks were used for visual cropping. SRR (M) without outlier rejection (OR) presents various artefacts. Similarly, the localization masks as used for SRR (L) lead to poor reconstruction outcomes despite the use of outlier rejection. The outlier-robust results SRR (M) and the proposed SRR (S) based on manual and automated brain masks, respectively, provide successful reconstructions and are, visually, almost indistinguishable. Green arrows indicate artefacts in SRR (M) without OR that are eliminated using our proposed OR method. Red arrows show differences between our proposed method and Kainz et al. (M).

Table 2.4.: Assessment of the robustness of the proposed template-space alignment approach. The comparison shows the number of successful template space alignments based on a total of 39 cases with 7 normal (group A), 16 pre-surgical and 16 post-surgical spina bifida cases (groups B1 and B2). FLIRT is based on correlation ratio as similarity measure whereas NIFTYREG uses symmetric block-matching based on NCC. Generally, NIFTYREG achieves a more robust alignment given a sufficiently good initial alignment. Using our proposed principal brain axes (PBA)-initialized block-matching registration step, a very robust template-space alignment without a failure case can be achieved even for pathological brains. The SRR (S) with the overlaid SRR (L)/(S)/(M) brain masks for the failed B1 case is shown in Figure 2.14.

	SRR (L)			SRR (S)			SRR (M)		
	A	B1	B2	A	B1	B2	A	B1	B2
NIFTYREG	0	0	0	1	0	0	1	0	0
FLIRT	0	0	0	4	0	4	3	0	4
FLIRT \circ PBA-init	0	0	0	7	0	4	4	0	3
NIFTYREG \circ PBA-init	2	0	0	7	15	15	7	16	16
Total number of cases	7	16	16	7	16	16	7	16	16

illustrates that all methods apart from SRR (S) produce statistically significant differences compared to SRR (M) in terms of measured slice similarities. Thus, SRR (S) and SRR (M) appear of similar volumetric self-consistency as quantified by the similarities between motion-corrected and respectively projected HR volume slices.

In absence of a ground-truth of the HR volume, an additional subjective quality assessment in a clinical context was made. Two pediatric radiologists (MA and PP) assessed all reconstructions in the template-space side-by-side blinded to the reconstruction methods. The HR masks obtained by SRR (M) were used for a consistent visual cropping of the reconstructions. The radiologists gave scores of three metrics on the results: 1) anatomical clarity of the cerebellar structure (CS), the cerebral aqueduct and the interhemispheric fissure (CAIF) and the longitudinal cerebral fissure (LCF) in the range of $[0, 4]$, 2) SRR quality against introduced artifactual structures and edge uncertainty in the range of $[0, 2]$, and 3) radiologists' preference in the range of $[0, 2]$. A higher score for each metric indicates a better reconstruction. The evaluation results are summarized in Figure 2.17 (a more detailed comparison of the individual scores is provided in the Figures A.2 and A.3). It shows that SRR (S) and SRR (M) achieved high-quality reconstruction results that are subjectively al-

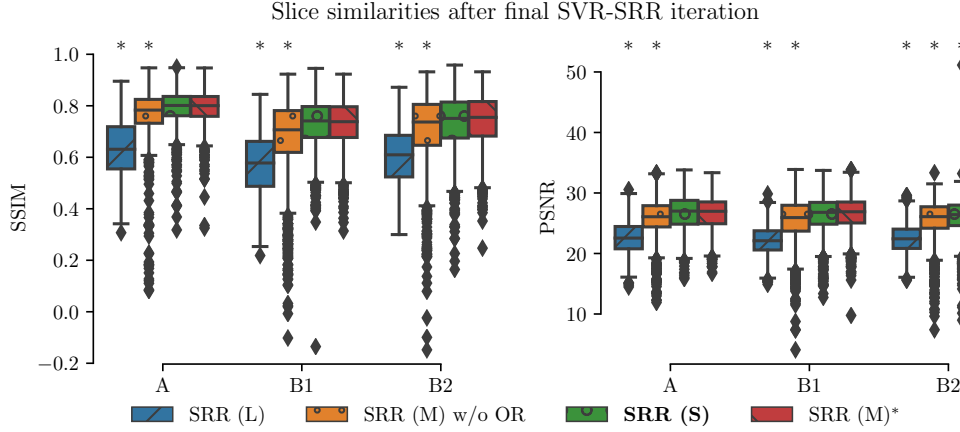


Figure 2.16.: Quantitative comparison of different reconstruction methods based on $\text{Sim}(\mathbf{y}_k^i, \mathbf{A}_k^i \mathbf{x}^i)$ after the final SVR-SRR iteration ($i = 3$) in terms of SSIM and PSNR. A * denotes a significant difference compared to SRR (M) within each group based on Kruskal-Wallis with post-hoc Dunn tests ($p < 0.05$). Thus, SRR (S) and SRR (M) appear of similar volumetric self-consistency as quantified by the similarities between motion-corrected and respectively projected HR volume slices.

most indistinguishable. Moreover, both SRR (S) and SRR (M) were consistently preferred over Kainz et al. (M) by the radiologists which can be explained by the high anatomical clarity and SRR quality achieved by our proposed reconstruction framework. The comparison against Kainz et al. (M) confirms the effectiveness of our proposed outlier-robust SRR framework which is also illustrated in Figures 2.15 and 2.18 to 2.22.

Using our non-optimized implementation on a single computer with four CPUs, the typical processing time for SRR (S) was approximately 13 min for the subject-space reconstruction, i.e. the computation of the two-step iterative motion-correction and volumetric reconstruction steps, and approximately 11 min for the template-space reconstruction, i.e. the combined template-space alignment and volumetric reconstruction from motion-corrected slices. The reconstruction times for SRR (M) were comparable. For Kainz et al. (M) the subject-space reconstructions took approximately 6 min on average.

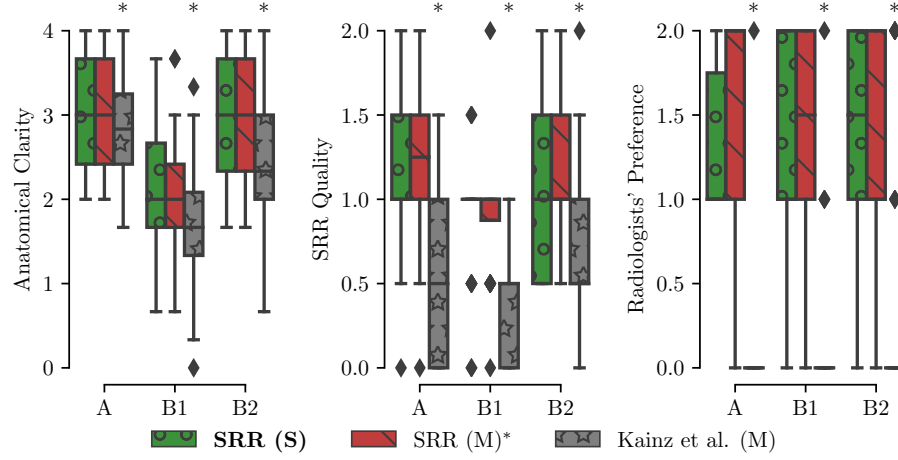


Figure 2.17.: Summary of clinical evaluation. Two radiologists performed a qualitative assessment of the obtained HR reconstructions regarding anatomical clarity, SRR quality and subjective preference involving 39 cases. A higher score indicates a better outcome. For anatomical clarity scores indicate how well CS, CAIF and LCF are visualized in each image with ratings 0 (structure not seen), 1 (poor depiction), 2 (suboptimal visualization; image not adequate for diagnostic purposes), 3 (clear visualization of structure but reduced tissue contrast; image-based diagnosis feasible), and 4 (excellent depiction; optimal for diagnostic purposes). SRR quality is a combined average score of individual visible artefacts and blur scores with ratings 0 (lots of artefacts/blur) to 2 (no artefact/blur). Radiologists' preference ranks subjectively from the least (0) to the most preferred (2) reconstruction. A * denotes a significant difference compared to SRR (M) based on a Wilcoxon signed-rank test ($p < 0.05$). The results underline that SRR (M)/(S) represent high-quality reconstructions with high anatomical clarity that are visually indistinguishable and were subjectively preferred over Kainz et al. (M) by the two radiologists.

2.4. Discussion

Our automated pipeline for fetal brain reconstruction in MR imaging benefits from deep learning-based localization and segmentation where a CNN-based coarse segmentation is proposed for robust localization and a multi-scale loss function for a fine segmentation of the fetal brain. Compared with [Keraudren et al., 2014], our localization method does not need prior information such as shape and size of the fetal brain and it achieved superior performance in less time. Different from [Salehi et al., 2018], which takes a whole image as the input of a CNN, our segmentation method follows a coarse-to-fine approach. The benefit is three-fold: 1) it rejects false

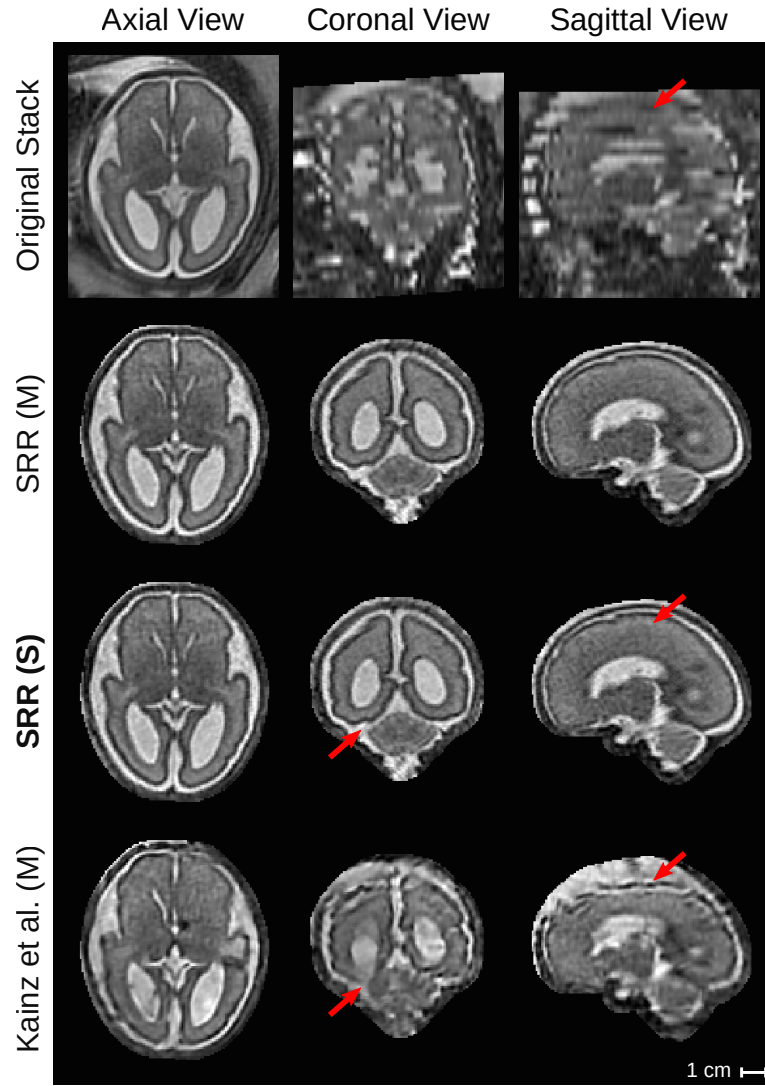


Figure 2.18.: Qualitative comparison of reconstruction methods in the template space. The comparison shows the template space reconstructions of a group B2 subject (post-surgical SB, GA = 27 weeks) based on 7 LR input stacks. An original stack (linearly resampled) with resolution of $0.47^2 \times 3 \text{ mm}^3$ is provided for reference. Red arrows show anatomical differences between SRR (S) and Kainz et al. (M).

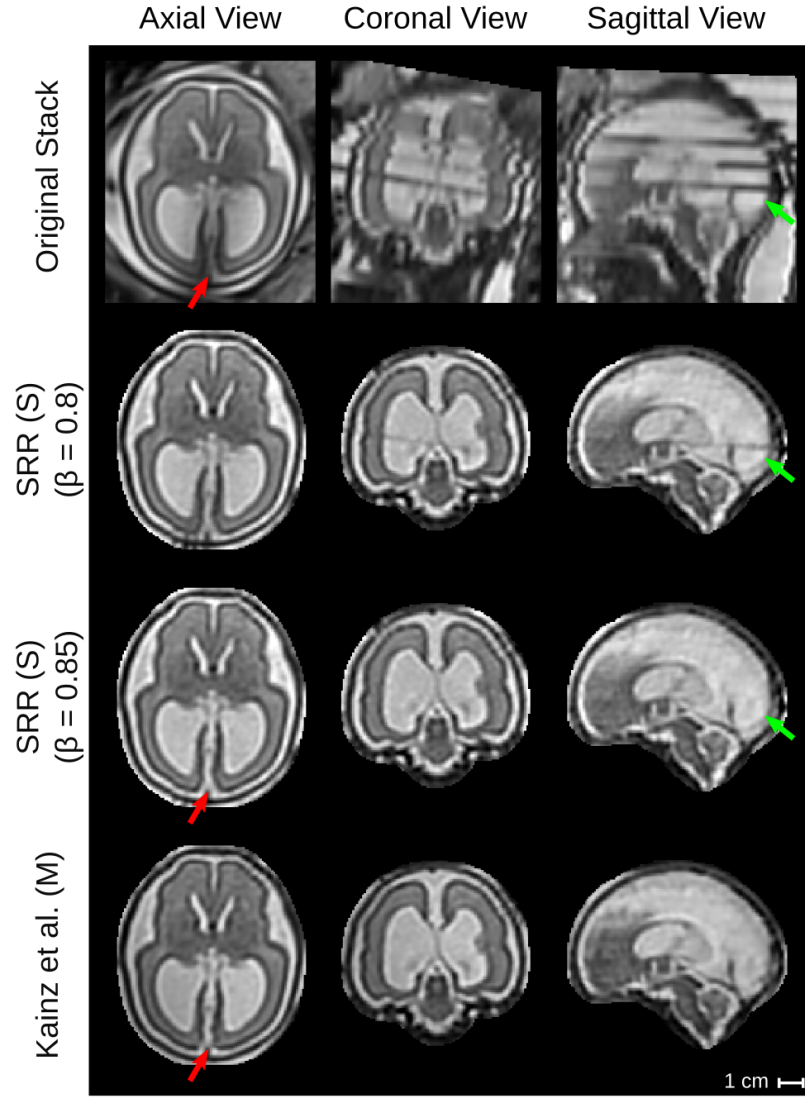


Figure 2.19.: Qualitative comparison of reconstruction methods in the template space. The comparison shows the template space reconstructions of a group B2 subject (post-surgical SB, GA = 26 weeks) based on 4 LR input stacks. An original stack (linearly resampled) with resolution of $0.74^2 \times 3 \text{ mm}^3$ is provided for reference. Green arrows indicate the rejection of the final intensity-artefacted slice of the original stack using the outlier-threshold $\beta = 0.85$. Red arrows show anatomical differences between SRR (S) and Kainz et al. (M) in direct comparison with the original stack.

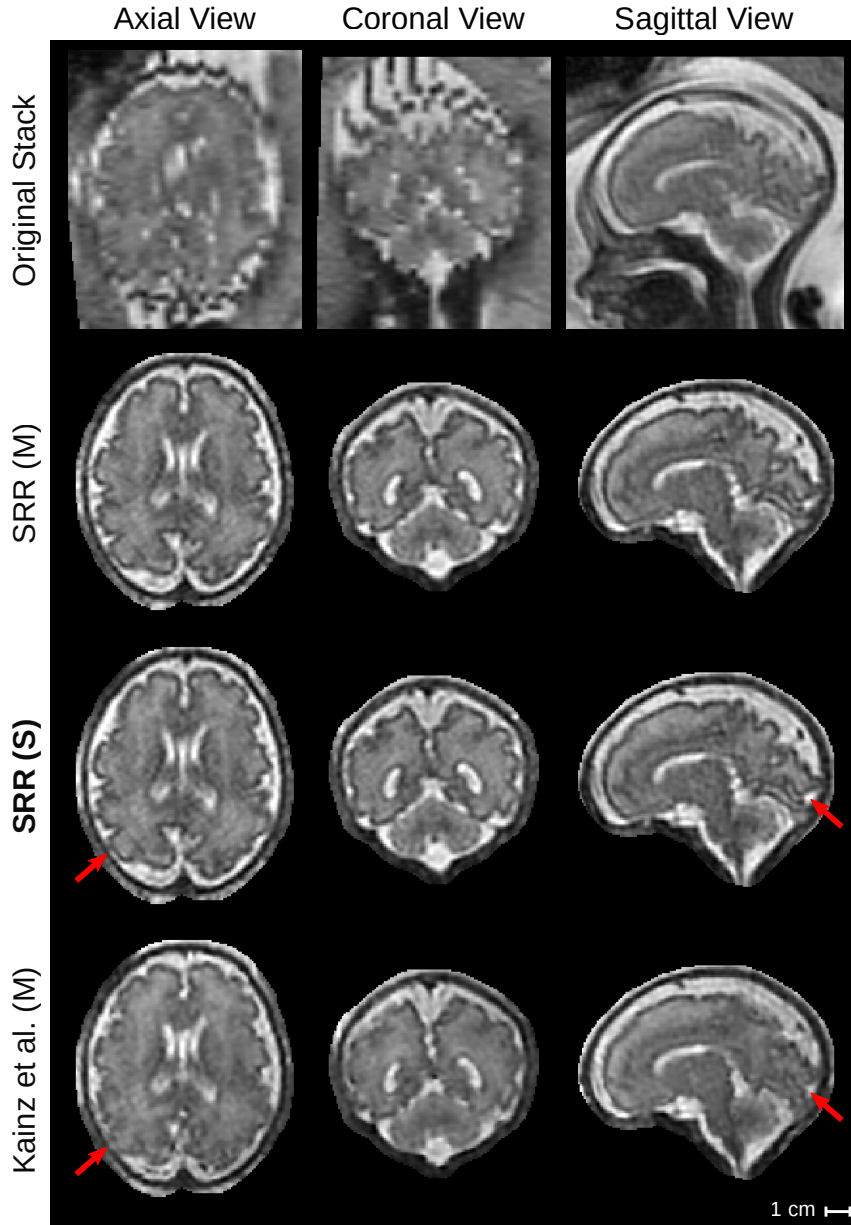


Figure 2.20.: Qualitative comparison of reconstruction methods in the template space. The comparison shows the template space reconstructions of a group A subject (normal, GA = 30 weeks) based on 7 LR input stacks (subject space SRRs are shown in Figure 2.15b). An original stack (linearly resampled) with resolution of $0.74^2 \times 3 \text{ mm}^3$ is provided for reference. Red arrows indicate example differences in the reconstruction outcomes compared to Kainz et al. (M).

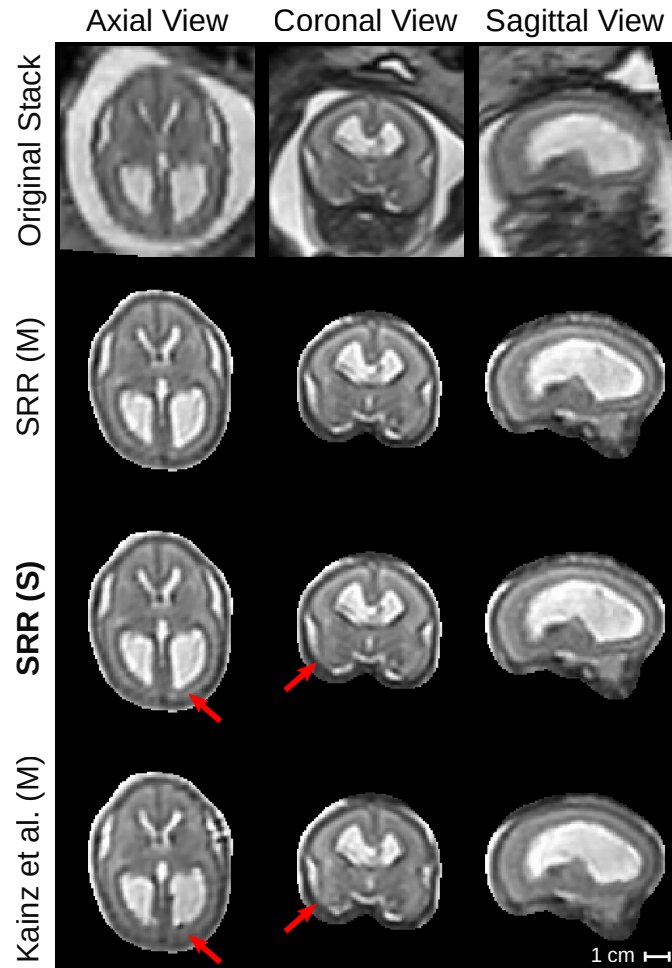


Figure 2.21.: Qualitative comparison of reconstruction methods in the template space. The comparison shows the template space reconstructions of a Group B1 subject (pre-surgical SB, GA = 23 weeks) based on 6 LR input stacks. An original stack (linearly resampled) with resolution of $0.74^2 \times 3 \text{ mm}^3$ is provided for reference. Red arrows show differences between SRR (S) and Kainz et al. (M).

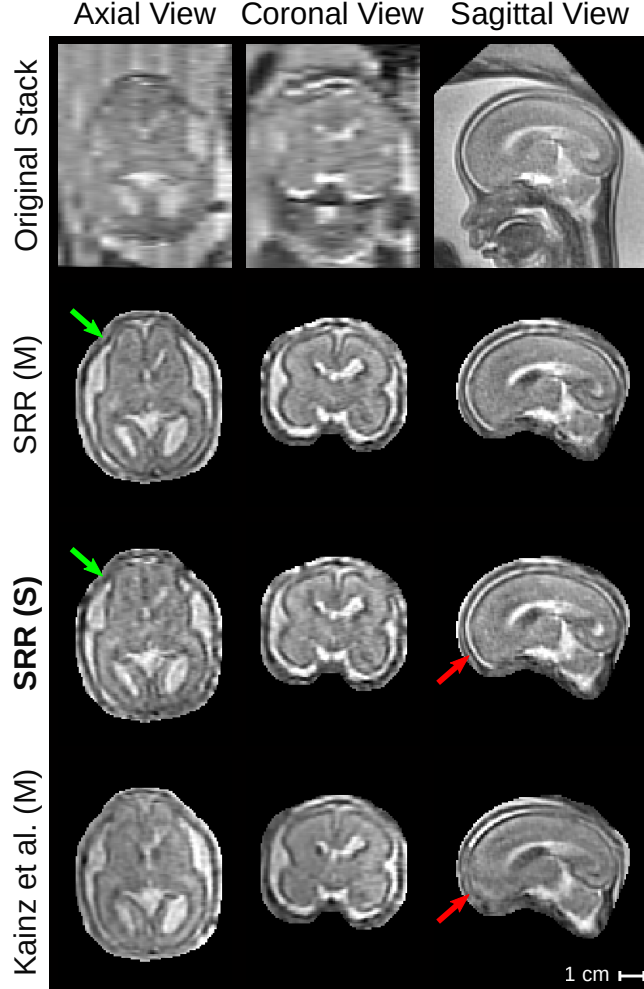


Figure 2.22.: Qualitative comparison of reconstruction methods in the template space. The comparison shows the template space reconstructions of a group B1 subject (pre-surgical SB, GA = 23 weeks) based on 6 LR input stacks. An original stack (linearly resampled) with resolution of $0.39^2 \times 4 \text{ mm}^3$ is provided for reference. It represents the only case where SRR (M) is markedly better than SRR (S). Green arrows indicate differences between SRR (M) and SRR (S). Red arrows show differences between SRR (S) and Kainz et al. (M).

positives outside the localization result, 2) the training data for Seg-Net are local regions around the fetal brain and therefore they have less imbalance between the foreground and background, and 3) Seg-Net has lower memory requirements and is more efficient than working on the whole image. However, it requires training of two networks independently, and may be improved by joint training or adopting attention mechanisms [Oktay et al., 2018, He et al., 2017] in the future. Minimizing our proposed multi-scale loss function encourages a segmentation to be close to the ground truth at multiple non-local levels, and helps to obtain more spatially consistent results as shown in Figure 2.9.

Moreover, we presented an alternative reconstruction framework that includes a novel outlier-rejection method for robust SRR. In contrast to [Gholipour et al., 2010b] and [Kuklisova-Murgasova et al., 2012], our formulation leads to a simple, yet effective, linear least-squares problem with a single hyperparameter that can be solved efficiently. Despite its simplicity, this outlier-rejection method using a single threshold parameter value was shown to allow for a remarkably robust elimination of outliers for most cases (Figure 2.19). We demonstrate that our proposed outlier-robust framework can produce high-quality HR visualizations from highly heterogeneous and challenging clinical data with results superior to the state-of-the-art toolkit Kainz et al. (M) [Kainz et al., 2015b, Kuklisova-Murgasova et al., 2012]. This involved stacks with multiple image resolutions per case including high slice thicknesses between 2.5 mm and 4.4 mm which can be severely affected by substantial motion and intensity artefacts. In particular, we show that high-fidelity reconstructions with clear tissue boundary definitions can be achieved even in case a corrupted target stack is selected. However, limitations of the comparison against Kainz et al. (M) include that the publicly available method²⁶ only allows to specify a single mask for the target stack whereas our method can take advantage of using separate masks for all input stacks. This allows a more accurate motion correction for our method and can therefore lead to higher quality HR reconstructions. A potential exclusion of heavy motion- or artefact-corrupted stacks is likely to improve the SRR

²⁶<https://github.com/bkainz/fetalReconstruction>

quality further for cases that performed less satisfying (e.g. Figure 2.22).

We investigated a novel robust motion correction method to encourage higher consistency of spatial transformations within each sub-stack of an interleaved stack acquisition. The approach based on Gaussian process regularization was shown to achieve higher retention rates of slices so that, ultimately, the volumetric reconstruction algorithm can take advantage of more input data due to fewer slice rejections. However, our preliminary results indicated that the currently proposed method achieves a minimal improvement only and does not lead to significantly higher slice retention rates. A more thorough investigation is needed, but based on the high reconstruction quality outcomes achieved by the “standard” SVR, rejections due to slice misregistrations do not appear to be a major concern in our proposed reconstruction framework.

Additionally, we present a fast template space alignment method for high-resolution visualizations in the standard anatomical planes that is robust to large brain morphology changes such as encountered in spina bifida. Further robustness to more substantial false-positives in the brain mask HR volume (Figure 2.14) could be achieved by using robust principal component analysis [Candès et al., 2011, Parikh and Boyd, 2014] to estimate the principal brain axes. Faster computational times for the HR volume reconstructions can be achieved by more efficient multi-core or GPU implementations including an, in principle, trivially parallelizable computation of the, currently, sequentially performed rigid slice-to-volume registrations.

In conclusion, we present a fully automated, and publicly available²⁷, framework for fetal brain MRI localization, segmentation and super-resolution reconstruction. Our experiments with fetuses with normal brain anatomy as well as fetuses with brain changes associated with spina bifida show that the proposed pipeline produces automatic reconstructions that are comparable to manual segmentation-based reconstructions, therefore, effectively eliminating the need of any manual intervention. In the future, we aim to apply this framework to quantify the impact of spina bifida

²⁷The automatic localization and segmentation framework FETAL_BRAIN_SEG is available at https://github.com/gift-surg/fetal_brain_seg and is integrated in the outlier-robust SRR framework NIFTYMIC (<https://github.com/gift-surg/NiftyMIC>).

surgical closure by measuring the resolution of the Chiari type II malformation and the degree of ventriculomegaly. Furthermore, applications of our method are not confined to spina bifida and may prove useful for a variety of other fetal brain pathologies where automatic high-resolution visualizations in the standard anatomical planes are clinically relevant.

CHAPTER 3.

Super-Resolution Reconstruction of Upper Abdominal MRI

Table of Contents

3.1. Introduction	107
3.2. Motion Correction and Volumetric Reconstruction for Upper Abdominal MRI	111
3.2.1. SRR using Reference-Guided Multimodal Deformable Mo- tion Correction	113
3.2.2. Outlier-robust SRR using Monomodal Rigid Motion Cor- rection	115
3.2.3. Point-Spread-Function-Aware Slice-to-Volume Registration	116
3.3. Point-Spread-Function-Aware Slice-to-Volume Registra- tion: Application to Upper Abdominal MRI Super- Resolution	117
3.3.1. Data, Evaluation Methodology and Results	118
3.3.2. Discussion	124
3.4. Super-Resolution for Upper Abdominal MRI: Acquisi- tion and Post-Processing Protocol Optimisation	125
3.4.1. Image Acquisition	126
3.4.2. Data, Evaluation Methodology and Results	129
3.4.3. Discussion	146

Foreword. This chapter is adapted and contains content from the work published in

- **Ebner, M.**, Chouhan, M., Patel, P. A., Atkinson, D., Amin, Z., Read, S., Punwani, S., Taylor, S., Vercauteren, T., and Ourselin, S. (2017a). Point-Spread-Function-Aware Slice-to-Volume Registration: Application to Upper Abdominal MRI Super-Resolution. In Zuluaga, M. A., Bhatia, K., Kainz, B., Moghari, M. H., and Pace, D. F., editors, *Reconstruction, Segmentation, and Analysis of Medical Images. RAMBO 2016*, volume 10129 of *Lecture Notes in Computer Science*, pages 3–13. Springer International Publishing
- **Ebner, M.**, Patel, P. A., Atkinson, D., Caselton, L., Firmin, L., Amin, Z., Bainbridge, A., De Coppi, P., Taylor, S. A., Ourselin, S., Chouhan, M. D., and Vercauteren, T. (2019a). Super-resolution for upper abdominal MRI: Acquisition and post-processing protocol optimization using brain MRI control data and expert reader validation. *Magnetic Resonance in Medicine*, page mrm.27852
- **Ebner, M.**, Patel, P. A., Atkinson, D., Caselton, L., Taylor, S. A., Bainbridge, A., Ourselin, S., Chouhan, M., and Vercauteren, T. (2019b). Reconstruction-based Super-Resolution for High-Resolution Abdominal MRI: A Preliminary Study. In *International Society for Magnetic Resonance in Medicine (ISMRM)*
- Chouhan, M., **Ebner, M.**, Patel, P. A., Atkinson, D., Firmin, L., Amin, Z., De Coppi, P., Ourselin, S., Vercauteren, T., and Taylor, S. A. (2019). Expert-reader validation of optimised Super-Resolution Reconstruction for upper abdominal MRI. In *European Society of Gastrointestinal and Abdominal Radiology (ESGAR)*.

My specific contributions presented in this chapter are:

- Introduction of super-resolution reconstruction methods to reconstruct an isotropic, high-resolution 3D visualization of upper gastrointestinal (GI) anatomy from motion-corrupted, low-resolution, single-shot T2-weighted (SST2W) MR cholangio-pancreatography (MRCP) sequences.

- Methodological contributions include:
 - Introduction of a non-iterative three-step volumetric reconstruction algorithm based on a reference-guided multimodal, in-plane deformable motion correction approach. This framework is specifically tailored to reconstruct GI anatomy from sequences that are routinely available in clinical MRCP protocols.
 - Presentation of a point spread function (PSF)-aware slice-to-volume registration method which takes into consideration the physical slice acquisition process
 - Design and execution of a pilot study involving the acquisition of multiplanar SST2W stacks of the upper abdomen and the brain ('quasi-static' control data, to remove the effect of upper abdominal motion artefact) for eight healthy volunteers with the overall objectives of:
 - * optimising source image acquisition protocols by establishing the ideal number and orientation of SST2W series for MRCP SRR generation
 - * optimising post-processing protocols by defining the most suitable approach to registration/motion correction for SRR in the upper abdomen
 - * validation of the overall potential of upper abdominal SRR using expert-readers

A follow-up work on a patient cohort is currently in preparation to assess the clinical utility of MRCP SRR obtained by the developed framework.

3.1. Introduction

As introduced in Chapter 1, Magnetic resonance cholangio-pancreatography (MRCP) is an established method for imaging the upper abdomen and biliary/pancreatic ducts (Figure 1.2). Heavily T2-weighted (HT2W) sequences exploit slow moving fluid in the biliary and pancreatic ducts to generate high-resolution images of the biliary tree [Barish et al., 1999, Griffin et al., 2012]. Such HT2W images are acquired as near-isotropic voxel three-dimensional (3D) image volumes during free breathing us-

ing respiratory triggers and are useful in the assessment of intra-ductal benign and malignant pathology [Basaran et al., 2008, Chen et al., 2015, Howard et al., 2006]. However, the assessment of peri-ductal and extra-ductal upper abdominal soft tissue pathology is reliant on traditional two-dimensional (2D) single-shot T2-weighted (SST2W) imaging (e.g. half-Fourier acquisition single-shot turbo spin echo, HASTE sequences) because of the more suitable tissue contrast they provide of surrounding anatomy. To achieve acceptable in-plane signal-to-noise ratio (SNR), slice thickness is increased, with resultant low through-plane resolution, anisotropic voxel 2D images. Because of the close proximity of fine ductal/vascular structures in the upper abdomen, these 2D images are particularly susceptible to partial voluming effects (PVEs), whereby signal from a single voxel is contaminated by signal from multiple anatomical structures. Images are also obtained in breath-hold, so that patient non-compliance and breath-hold difficulties commonly introduce inter-slice motion artefact. To mitigate these effects, SST2W sequences can be obtained consecutively in axial and coronal planes, with radiologists reading low through-plane resolution, motion-artefacted image series in both planes to improve sensitivity to pathology. However, early malignant lesions are typically small and mural/extra-ductal (rather than intra-ductal) and easily overlooked. Diagnostic pathways are therefore reliant on non-MR imaging modalities for the exclusion of small volume pathology, but these can be invasive (e.g. endoscopic ultrasound, EUS) or require ionising radiation (e.g. computed tomography, CT).

Chapter 1 presented Super-Resolution Reconstruction (SRR) as a post-processing technique to combine multiple low-resolution (LR) 2D image stacks into a single high-resolution (HR), 3D visualization. Applications of SRR in MR imaging (MRI) range from adult studies on the tongue [Woo et al., 2012] and thorax [Van Reeth et al., 2015] to fetal applications [Rousseau et al., 2010, Kainz et al., 2015b, Ebner et al., 2018c] (Chapter 2). Despite being well-suited to overcome the limitations of multiplanar SST2W in principle, its application in the upper abdomen to-date has been limited. As pointed out in Chapter 1, Super Resolution (SR) can only work accurately in case of very precise motion estimation with subvoxel accuracy for all LR observations for

the recovery of subvoxel detail [Park et al., 2003, Van Reeth et al., 2012, Milanfar, 2010]. This is especially difficult in the context of abdominal imaging where images acquired from separate breath-holds are subject to inspiratory/expiratory variation in addition to deformation arising from cardiac motion, arterial pulsation and gastrointestinal tract peristaltic motion. Existing respiratory motion models for motion correction require the availability of respiratory surrogate data [McClelland et al., 2013] which are currently not available for MRCP studies. In cardiac MRI, motion-correction approaches typically rely on the assumption that images are acquired at known cardiac phases [Villard et al., 2017]. Using an SRR approach such as the iterative two-step registration-reconstruction approach used in fetal MRI [Kainz et al., 2015b, Rousseau et al., 2006], applied to only two stacks, is prone to generate a strongly biased volume and the currently used rigid motion models might not be sufficient.

In the preliminary work of our first study [Ebner et al., 2017a], presented in Section 3.3, we were the first to apply Super-Resolution (SR) to MRCP studies and introduced an alternative SRR framework specifically tailored to reconstruct upper gastrointestinal (GI) anatomy for sequences that are routinely available in clinical protocols. Guided by the HT2W volume as a 3D reference, a rigid slice-to-volume registration (SVR) was performed followed by an in-plane deformation step for each individual slice to compensate for non-rigid deformations. In this preliminary study, we identified several key challenges:

1. *Optimizing imaging planes is required for adequate upper GI SRR:* Super-resolution applied to one axial and one coronal SST2W stack allows for a 3D visualization of peri-ductal anatomy in higher resolution. However, datasets with three or more planes would be justifiable in routine clinical imaging protocols if they can be used to generate a higher diagnostic yield SRR due to improved PVE correction [Rousseau et al., 2010, Shilling et al., 2009]. However, the theoretical insight on the optimal orientation and number of input stacks for SRR approaches is relatively limited [Baker and Kanade, 2002, Lin and Shum, 2004, Candès and Fernandez-Granda, 2014]. This is especially the

case in abdominal applications where motion patterns can be relatively complex and magnification factors for SR are very high given the large disparity between through-plane vs inplane resolutions.

2. *Optimizing motion-correction strategies for MRCP SRR:* HT2W-guided motion-correction led to promising SRR results for a relatively small region of interest (ROI) around biliary/pancreatic ducts. However, for clinically useful HR visualizations, larger ROIs showing the entire biliary tree anatomy will need to be reconstructed. As peri-ductal anatomy is not demonstrated in HT2W volumes, different motion-correction references or strategies might be more suitable for this task. Moreover, the bigger the ROI, the more critical it is to correct for the complex, non-rigid motion presented in multiple SST2W stacks prior to the SRR step.

We hypothesized that the application of SRR algorithms to a large numbers of retrospective MRCP scans potentially available is unlikely to address these. Instead, the prospective evaluation of SRR in a clinical setting will require the introduction of changes to the established clinical MRCP protocol. The rationale for the second study [Ebner et al., 2019a], presented in Section 3.4, was to collect data in normal volunteers and optimise the technique so that clinical MRCP protocols can then be amended to minimise patient inconvenience and maximise the likelihood of successful SRRs in patients. We therefore designed and conducted a pilot study with the following contributions:

1. MRCP sequence optimisation for abdominal reconstruction-based SR reconstruction based on eight healthy volunteers:
 - Dense T2-weighted ultra-fast 2D MR image sampling using SST2W imaging in multiple imaging planes including seven different image orientations:
 - standard anatomical acquisitions, i.e. axial, coronal and sagittal images
 - additional standard anatomical acquisitions in axial, coronal and sagittal orientation with an offset of half the slice thickness in the respec-

tive slice-select direction

– four oblique orientations

- A non-standard balanced fast field echo (BFFE) 3D sequence is acquired in addition to the standard-clinical HT2W volume. As it lacks peri-ductal tissue contrast it is not clinically useful in the context of MRCP diagnosis but due to its high resolution and similar T2W contrast it may represent a suitable reference for image-guided motion-correction strategies.
 - The identical abdominal protocol is used for imaging the brain but with an additional HR T2W volume invaluable for method optimisation and validation in a controlled, quasi-static scenario.
2. Experimental estimation of optimal number of SST2W stacks and their orientation to trade off SRR quality and acquisition time specific for MRCP studies
 3. Qualitative and quantitative assessment of various motion correction approaches for MRCP studies to obtain best SRR outcomes. This includes the assessment of
 - reference-guided multimodal¹ registration approach for SST2W images guided by a reference image of a different modality [Ebner et al., 2017a].
 - suitability of monomodal rigid motion-correction and robust reconstruction approaches [Kainz et al., 2015b, Ebner et al., 2018c] based on SST2W image information only.

3.2. Motion Correction and Volumetric Reconstruction for Upper Abdominal MRI

Assuming a classical slice acquisition model [Greenspan, 2008, Gholipour et al., 2010b] for each LR 2D SST2W slice acquisition $\mathbf{y}_{s,i} \in \mathbb{R}^{N_s}$ from a stack $s \in \mathbf{S}$ with

¹With *multimodal* registration, we refer to the involvement of two images with different MR image contrasts due to different acquisition protocols as opposed to *monomodal* registration where two images of the same image contrast are used.

slice index $i \in \mathbf{I}_s$, the pixelwise association with the unknown HR volume $\mathbf{x} \in \mathbb{R}^N$, whereby $N_s \ll N$ for the voxel numbers due to the LR 2D image acquisition, (1.1) can be rephrased as

$$\mathbf{y}_{s,i}(j) = A_{s,i}(j, \mathbf{x}) + \mathbf{n}_{s,i}(j) \in \mathbb{R} \quad \text{for all slice voxels } j = 1, \dots, N_s. \quad (3.1)$$

The linear operator $A_{s,i}(j, \cdot)$ acts as point spread function (PSF)-defined intensity interpolator in the HR volume space that approximates the image acquisition process at a non-linearly transformed physical position of voxel j of slice $\mathbf{y}_{s,i}$ up to the imaging noise $\mathbf{n}_{s,i}$. Each voxel intensity of a LR slice is therefore influenced by a certain neighbourhood of this voxel within a HR volume \mathbf{x} given by the assumed PSF that is specific to the slice profile of the LR MR acquisition [Liang and Lauterbur, 2000]. For SST2W sequences, Chapter 2 highlighted that a common approximation is given by a slice-aligned 3D Gaussian function that depends on the in- and through-plane resolution of the LR slice [Jiang et al., 2007, Kuklisova-Murgasova et al., 2012]. However, the motion that each anatomical region experiences during acquisition time is unknown. Once estimated, the HR volume can be obtained from the (motion-corrected) LR slices by solving the associated super-resolution reconstruction problem using a maximum a posteriori formulation (MAP) (1.2) [Van Reeth et al., 2012, Gholipour et al., 2010b, Ebner et al., 2017a]

$$\mathbf{x}^* := \arg \min_{\mathbf{x} \geq 0} \left(\sum_{s \in \mathbf{S}} \sum_{i \in \mathbf{I}_s} \frac{1}{2} \|\mathbf{y}_{s,i} - \mathbf{A}_{s,i} \mathbf{x}\|_{\ell^2}^2 + \frac{\alpha}{2} \|\nabla \mathbf{x}\|_{\ell^2}^2 \right) \in \mathbb{R}^N \quad (3.2)$$

where $\mathbf{A}_{s,i} \mathbf{x}$ denotes the application of (3.1) to a vector in \mathbb{R}^{N_s} , $\alpha \geq 0$ the regularization parameter and ∇ the differential operator. In particular, the linear operator $\mathbf{A}_{s,i} : \mathbb{R}^N \rightarrow \mathbb{R}^{N_s}$, $\mathbf{x} \mapsto \mathbf{A}_{s,i} \mathbf{x} =: \tilde{\mathbf{y}}_{s,i}$ models the image acquisition process that generates a LR slice $\tilde{\mathbf{y}}_{s,i}$ from a HR volume \mathbf{x} at a specific, and in our case, estimated, position and orientation within that HR volume. The second term of (3.2) is a first-order Tikhonov regularization which corresponds to a MAP formulation exploiting a probabilistic prior on the HR volume. This counteracts the ill-posed nature of the minimisation problem and retains a computationally efficient least-square structure.

The final HR volume \mathbf{x}^* is also referred to as the Super-Resolution Reconstruction (SRR). More complex SRR models have been proposed in addition to the MAP formulation including modifying (3.2) to rely on robust M-estimator [Gholipour et al., 2010b] and total variation formulations [Tourbier et al., 2015, Bastý et al., 2017]. However, while they substantially increase the computational cost, in our experience, they tend to show little improvement in the obtained reconstruction quality in case of appropriate motion correction. Associated experiments comparing Tikhonov and total variation regularizations are summarized in Section 3.4.2.

Two different motion-correction approaches were investigated for upper abdominal MRI reconstruction:

- (a) a multimodal reference-guided, in-plane deformable registration approach that registers the LR SST2W slices to a separately acquired 3D HR reference volume of a different contrast (e.g. HT2W or BFFE volumes)
- (b) a monomodal rigid motion-correction approach based on robust outlier rejection using only the SST2W slice image information

3.2.1. SRR using Reference-Guided Multimodal Deformable Motion Correction

In our preliminary study [Ebner et al., 2017a] (Section 3.3), we introduced a reference-guided motion-correction approach whereby the following is assumed:

- (a) the resolution of the reference image is sufficiently high to act as a 3D reference volume
- (b) the occurring anatomical deformation can be captured by deforming the slice only in the in-plane direction; the contribution in the orthogonal slice-select direction can therefore be neglected given the thick slices and the associated intensity information uncertainty due to PVEs.

Based on those assumptions the following non-iterative three-step volumetric reconstruction algorithm is performed (Figure 3.1):

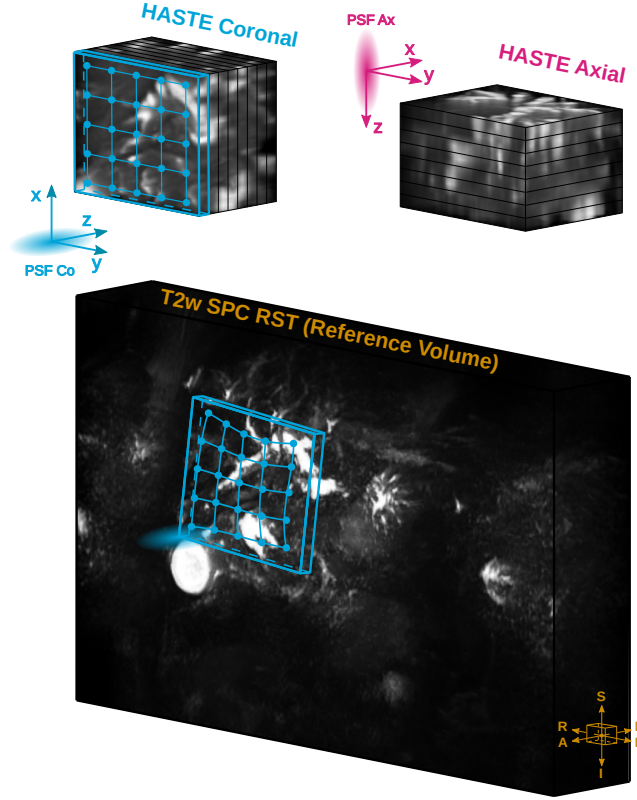


Figure 3.1.: Reference-guided motion correction based on rigid slice-to-volume registration and in-plane deformation steps (Section 3.2.1). Slices are rigidly motion corrected to find the best rigid motion estimate within the 3D reference before the non-rigid deformation step is applied. In-plane deformation is performed using a 2D non-rigid registration for the corresponding HASTE (fixed image) and intersecting reference (moving image) slices. Additional PSF-awareness for the rigid and in-plane non-rigid registration steps can be applied to take into consideration the physical slice acquisition process (Section 3.2.3).

- (i) multimodal slice-to-volume registration where each individual 2D slice of each stack is rigidly registered to the 3D reference
- (ii) based on the intersection of the slices with the 3D reference, each slice is deformed in-plane to compensate for non-rigid deformations
- (iii) estimation of the SRR volume by solving (3.2) using the estimated deformations

3.2.2. Outlier-robust SRR using Monomodal Rigid Motion Correction

Outlier-robust SRR using rigid motion correction has been proposed recently for fetal MRI and exploits data from series obtained in at least three orientations [Rousseau et al., 2006, Gholipour et al., 2010b, Kuklisova-Murgasova et al., 2012, Kainz et al., 2015b]. Using an iterative two-step registration-reconstruction approach, a volumetric reconstruction step is followed by a rigid SVR step until convergence using the SST2W image series only. ‘Outlier’ slices are detected during the SRR steps and rejected to prevent misregistered or artefact-corrupted image slices from affecting the final SRR outcome. In this study, we deploy our recently presented method for fetal brain MRI [Ebner et al., 2018c] (Section 2.2.3). This computes the volumetric reconstruction using an SRR formulation similar to (3.2), i.e.

$$\mathbf{x}^{k+1} := \arg \min_{\mathbf{x} \geq 0} \left(\sum_{s \in \mathbf{S}} \sum_{i \in \mathbf{I}_{s,\beta}^k} \frac{1}{2} \|\mathbf{y}_{s,i} - \mathbf{A}_{s,i}^k \mathbf{x}\|_{\ell^2}^2 + \frac{\alpha}{2} \|\nabla \mathbf{x}\|_{\ell^2}^2 \right) \quad (3.3)$$

for a slice-index set $\mathbf{I}_{s,\beta}^k := \left\{ i : \text{Sim}(\mathbf{y}_{s,i}, \mathbf{A}_{s,i}^k \mathbf{x}^{k-1}) \geq \beta \right\} \subset \mathbf{I}_s$ containing only slices that demonstrate high agreement with their simulated counterparts projected from the previous SRR iterate using a similarity measure Sim and parameter $\beta > 0$. In particular, (3.2) is a special case of (3.3) since the combined linear forward operator $\mathbf{A}_{s,i}^k$ describes the oriented PSF-interpolator estimated after k motion-correction steps whereby only a subset of indices $\mathbf{I}_{s,\beta}^k \subset \mathbf{I}_s$ per stack s is considered. Thus, (3.3) represents a convex SRR problem with complete outlier removal using a single hyperparameter β in a linear least-squares minimisation formulation that can be solved efficiently using *matrix-free* operations [Diamond and Boyd, 2015].

Therefore, by using this iterative SRR framework it is assumed that

- (i) sufficient input SST2W images are available to allow anatomically plausible reconstructions from these LR image stacks
- (ii) the anatomical motion captured per slice is approximately rigid for the specified region of interest (ROI) for most of the slices

- (iii) the outlier-rejection algorithm can reliably reject individual slices that present non-matching deformations

3.2.3. Point-Spread-Function-Aware Slice-to-Volume Registration

Current motion correction techniques do not take into account the PSF for registration. This is particularly problematic since neglecting the PSF during resampling introduces aliasing and subsequently results in additional loss of information [Cardoso et al., 2015, Chacko et al., 2015]. The intent of using a PSF-aware registration is to blur the moving image (3D reference) with the PSF defined by the relative position between fixed image (LR 2D slice) and moving image in order to make them comparable during the registration process which may increase the accuracy for the motion-correction estimate [Cardoso et al., 2015, Chacko et al., 2015]. The “PSF-aware” registration model itself follows by the definition of $A_{s,i}(j, \cdot)$ as PSF-defined intensity interpolator in the HR volume space that approximates the image acquisition process at a non-linearly transformed physical position of voxel j of slice $\mathbf{y}_{s,i}$ in (3.1). In practice, a reasonable approximation for single-shot sequences in the slice-coordinate system has been found to be a 3D Gaussian defined by the variance-covariance matrix $\tilde{\mathbf{B}}_{s,i} := \text{diag}\left(\frac{(1.2s_1)^2}{8\ln(2)}, \frac{(1.2s_2)^2}{8\ln(2)}, \frac{s_3^2}{8\ln(2)}\right)$ with s_1, s_2 being the spacing in-plane and s_3 through-plane [Cardoso et al., 2015, Jiang et al., 2007] of slice $\mathbf{y}_{s,i}$. For the registration, this variance-covariance matrix needs to be expressed in the coordinate system of the moving image in order to accommodate the interpolation in the moving space accordingly. By accounting for the orientation difference between the LR slice and the HR volume with a basis transform with orthogonal matrix $\mathbf{U}_{s,i}$, the PSF is expressed by $\mathbf{U}_{s,i}^T \tilde{\mathbf{B}}_{s,i} \mathbf{U}_{s,i}$ for each single point with respect to the floating space. Thus, a PSF-aware slice-to-volume registration can be implemented by providing an *oriented Gaussian interpolator* $A_{s,i}$ for each slice $\mathbf{y}_{s,i}$ to a generic registration framework which updates the PSF depending on the current transformation parameters. The operation $A_{s,i}(j, \cdot)$ can be efficiently computed as a matrix-vector multiplication without storing a matrix explicitly by iterating over the $N_s \ll N$ voxels in a multi-threaded fashion while considering the

oriented Gaussian-weighted 3D reference volume voxel intensities. An illustration of the PSF-aware registration algorithm that takes into consideration the physical slice acquisition process for motion correction is provided in Figure 3.1. PSF-aware in-plane deformation can be achieved by using an oriented Gaussian interpolator to compute the intersecting reference slice as acting moving image for a subsequent in-plane (2D) non-rigid registration.

3.3. Point-Spread-Function-Aware Slice-to-Volume Registration: Application to Upper Abdominal MRI Super-Resolution

In our preliminary study [Ebner et al., 2017a], we were the first to introduce SR for upper GI anatomy using currently available clinical data as shown in Figure 1.2.

Axial and coronal stacks are acquired consecutively and cannot be regarded as motion-free given the non-periodic respiratory motion, tissue deformation due to cardiac motion and arterial pulsation, peristaltic and other complex motion affecting the upper gastrointestinal anatomy. Accurate registration and reconstruction are complicated by the fact that in current clinical protocols usually only two single-shot T2-weighted stacks are available (in axial and coronal planes) with a slice thickness approximately five times higher than the in-plane resolution. Existing respiratory motion models require the availability of respiratory surrogate data [McClelland et al., 2013] which are currently not available for MRCP studies. Using an SRR approach such as the iterative two-step registration-reconstruction approach used in fetal MRI [Kainz et al., 2015b, Rousseau et al., 2006], applied to only two stacks, is prone to generate a strongly biased volume and the currently used rigid motion models might not be sufficient.

Thus, we presented the following contributions:

- (i) introduction of a novel PSF-aware slice-to-volume registration (SVR) method which takes into consideration the physical slice acquisition process (Section 3.2.3)

- (ii) novel use of an existing heavily T2-weighted volume available in MRCP studies to guide registration (Section 3.2.1)
- (iii) use of a novel SRR framework to reconstruct upper abdominal MRI using a single, consistent model to incorporate the PSF in both registration and reconstruction steps

3.3.1. Data, Evaluation Methodology and Results

Data and Data Preprocessing

MRCP studies of four anonymized patients, scanned at the University College London Hospital, London, were used for this study. Among the clinically acquired scans for MRCP studies, a set of axial and coronal 2D HASTE sequences and a 3D heavily T2-weighted SPC RST volume acquisition were performed, as shown in Figure 1.2. The acquisition parameters for the coronal stack were $TE = 91$ ms, $TR = 1350$ ms, flip angle of 170° with resolution of $1.25 \text{ mm} \times 1.25 \text{ mm} \times 6 \text{ mm}$. The respective parameters for the axial stack were $TE = 91$ ms, $TR = 1200$ ms, flip angle of 160° with resolution of $1.48 \text{ mm} \times 1.48 \text{ mm} \times 5.50 \text{ mm}$. The heavily T2-weighted volume was acquired in coronal direction with dimensions of $1.09 \text{ mm} \times 1.09 \text{ mm} \times 1.30 \text{ mm}$. HASTE images were preprocessed via an ITK bias field correction filter step [Tustison et al., 2010]. Rectangular masks were provided for both axial and coronal stacks to mark a region of clinical interest.

Parametrization of Reconstruction Pipeline

First, both axial and coronal stacks were rigidly aligned to the reference volume using normalized mutual information (NMI) as similarity measure. Similar to the PSF-aware SVR approach, the multimodal volume-to-volume registration was made PSF-aware by blurring the 3D reference with an oriented Gaussian filter considering the PSF defined by slice-select direction and slice dimensions for each stack. Then, the NMI-guided rigid slice-to-volume PSF-aware registration approach was deployed. The PSF-aware in-plane deformation was performed with the NIFTYREG software

using a fast free-form deformation algorithm [Modat et al., 2010] with normalized cross-correlation as similarity measure. The corresponding 3D reference (floating image) slice for the in-plane registration was computed by using the oriented Gaussian interpolator to estimate the intersecting reference slice. An illustration of the reference-guided PSF-aware motion correction steps is provided in Figure 3.1. Prior to the volumetric reconstruction a linear model was used to account for the intensity differences in the HASTE sequences due their different acquisition parameters. In order to initialize the SRR solver with a regular grid volume from motion corrected slices, we used a discrete variant of Nadaraya-Watson kernel regression as an efficient scattered data approximation (SDA) scheme for irregularly sampled inputs [Vercauteren et al., 2006]. It is based on nearest neighbour sampling onto a regular grid followed by a subsequent Gaussian blurring operation for each single slice. The initial value was computed by the SDA approach with $\sigma = 4$ to avoid the problem of inpainting during SRR. The corresponding SRR step was performed with the regularization parameter $\alpha = 0.03$ selected via L-curve studies. In this application, we chose a L-BFGS-B algorithm to deal with this large linear system (3.2) and its positivity constraints to solve the corresponding normal equations.

Evaluation Methodology

The algorithm was run with and without the consideration of the oriented PSF for all registration steps (PSF_0 or PSF_1) and with and without usage of the in-plane deformation model (DM_0 or DM_1) resulting in four different reconstructions for analysis. The reconstructions were initially quantitatively assessed by evaluating the residual via a normalized cross correlation metric, instead of the ℓ^2 -norm, in order to be insensitive to the intensity normalization step used to compensate for the different acquisition protocols. Following this, subjective assessment in a clinical context was made including direct comparison to reconstructions obtained by open-source toolkits successfully employed in the challenging problem of fetal MRI reconstructions (BTK-toolkit [Rousseau et al., 2013], version from 6 Jan 2016, and the IRTK-based toolkit [Kainz et al., 2015b], version from 11 Jun 2015). Two radiologists, blinded to

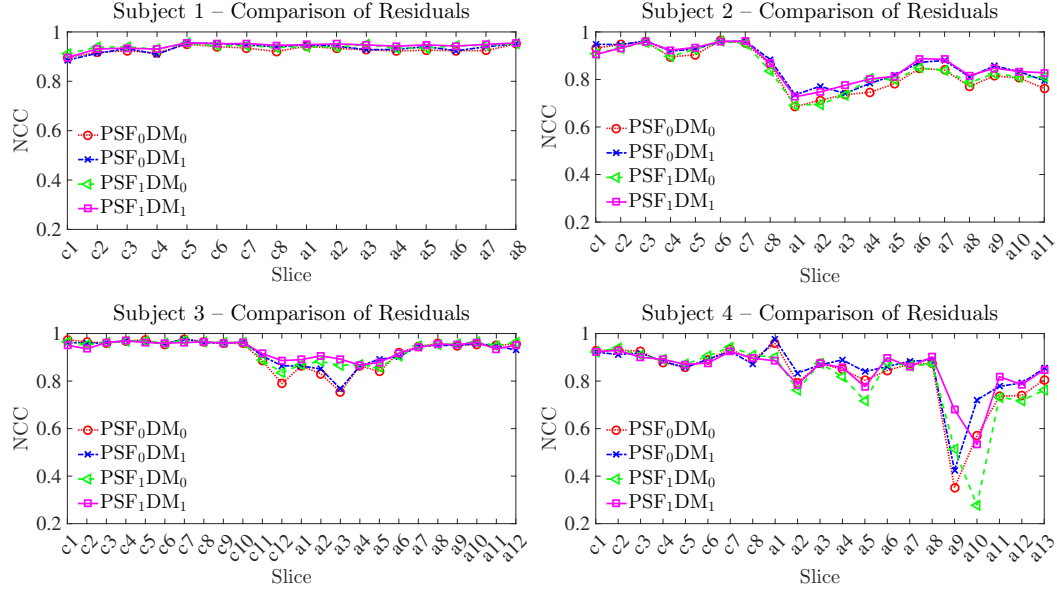


Figure 3.2.: Evaluation of the residuals for all subjects and modes of our proposed reconstruction framework visualized for all axial slices (a) and coronal slices (c). The associated NCC mean and standard deviation over all subjects for each mode are 0.88 ± 0.10 for PSF_0DM_0 , 0.89 ± 0.08 for PSF_0DM_1 , 0.88 ± 0.11 for PSF_1DM_0 and 0.90 ± 0.08 for PSF_1DM_1 , respectively.

the reconstruction methods, individually assessed reconstruction side-by-side and in comparison to the original HASTE data. The final score is a joint agreement of the radiologists' individual results. Scores were given for:

1. Clinical usefulness: based on how well common bile duct (CBD), left and right hepatic duct (LHD & RHD) were visualized and the degree of visible motion artifacts
2. Reconstruction quality: inferred by assessment of preservation of original structural information and the amount of additionally introduced artifactual structures
3. Radiologists' preferred reconstruction

Results

The evaluation of the residuals in (3.1) for all four subjects are visualized in Figure 3.2. The best agreement between the observed slice \mathbf{y}_k and simulated slice

$\mathbf{M}_k \mathbf{A}_k \mathbf{x}$ was obtained for the reconstruction which used the most comprehensive model including PSF-aware registration and in-plane deformable model (PSF₁DM₁). This is confirmed by calculating the mean of the residuals which rank PSF₁DM₁ ahead of all other variants. PSF₁DM₁ yields consistently better agreement for subjects 3 and 4 compared to other approaches which show less accurate registration results for some slices.

Table 3.1.: Summary of clinical evaluation averaged over all four subjects.

Evaluation included original HASTE data, four modes of our proposed reconstruction framework and reconstructions by other toolkits (BTK, IRTK). Clarity of anatomical structure score indicates how well CBD, LHD and RHD are visualized in each image with ratings 0 (structure not seen), 1 (poor depiction), 2 (suboptimal visualization; image not adequate for diagnostic purposes), 3 (clear visualization of structure but reduced tissue contrast; image-based diagnosis feasible) and 4 (excellent depiction; optimal for diagnostic purposes). Visible motion score rates the amount of visible non-corrected motion from score 0 (complete motion) to 3 (no motion). Preserved structural information score indicates how well original HASTE data information has been preserved with grades 0 (structures not identified), 1 (poor visualization of structures), 2 (clear visualization but not as good as originals) and 3 (as good as original). Introduced artifacts score rates the amount of additional artifactual structures from 0 (lots of new artifacts) to 2 (no new artifact). Radiologists' preference ranks the subjectively preferred reconstructions from 1 (least preferred) to 6 (most preferred) reconstruction.

	Clinical Usefulness		Reconstruction Quality		Radiologists' Preference
	Clarity of Anatomical Structures	Visible Motion	Preserved Structural Information	Introduced Artifacts	
HASTE	2.9 ± 0.3	1.8 ± 0.5	—	—	—
Ax & Cor					
PSF₀DM₀	2.9 ± 0.3	2.8 ± 0.5	2.0 ± 0.0	0.8 ± 0.5	4.2 ± 0.9
PSF₀DM₁	2.9 ± 0.3	2.5 ± 0.5	1.8 ± 0.5	1.0 ± 0.0	5.5 ± 1.0
PSF₁DM₀	2.9 ± 0.3	2.8 ± 0.5	2.0 ± 0.0	0.5 ± 0.5	3.5 ± 1.0
PSF₁DM₁	2.9 ± 0.3	2.8 ± 0.5	2.0 ± 0.0	0.5 ± 0.5	4.5 ± 0.5
IRTK	2.4 ± 0.5	2.8 ± 0.5	1.2 ± 0.1	0.0 ± 0.0	1.8 ± 0.5
BTK	1.9 ± 0.3	2.0 ± 0.0	1.0 ± 0.0	1.0 ± 0.0	1.2 ± 0.5

The radiologists' evaluation, shown in Table 3.1, indicates that the blinded radiologists had a clear preference for our novel PSF-aware SVR reconstructions. Additionally, our proposed reconstruction framework yields reconstructions of similar clarity of CBD, LHD and RHD as the original HASTE data. The reconstructions obtained via IRTK score slightly lower and it was felt that the images would not be suitable for making a clinical diagnosis. Furthermore, all reconstruction approaches demon-

strate their ability to correct for motion visible in the HASTE data. With regards to preservation of information in the original HASTE stacks, our novel PSF-aware SVR reconstructions are close to the originals' whereas IRTK and BTK² perform less satisfactorily. All reconstruction methods, to some degree, introduce structures which cannot be directly visualized by the original HASTE data.

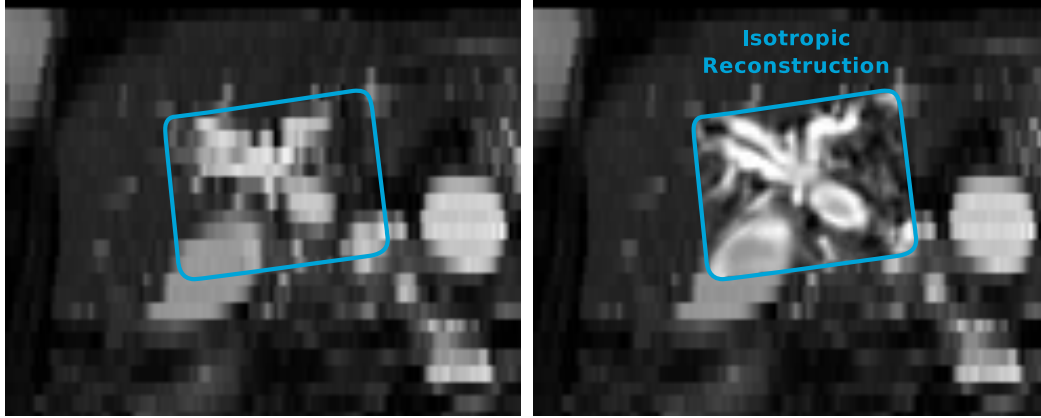


Figure 3.3.: Visualization of original HASTE data and the field of view for the SR reconstruction algorithms. The axially acquired HASTE stack is visualized in the through-plane direction showing the limited resolution due to the thick-slice acquisition (left). The isotropic, HR reconstruction (right) obtained by combining both axial and coronal HASTE stacks with the proposed uniform PSF-aware SRR algorithm is overlaid for direct comparison.

Figure 3.3 shows the reconstructed region obtained by the proposed framework in direct comparison with the original data. In Figure 3.4 our reconstruction variant PSF_1DM_1 and the reconstructions based on IRTK and BTK of one subject are provided along with the linearly resampled original data for comparison. This demonstrates that our proposed reconstruction framework largely preserves axial and coronal HASTE data information with minor degradation in image quality as opposed to both IRTK and BTK reconstructions. Moreover, it reveals sharp tissue delineation also in sagittal section where no image stack information is provided.

²The BTK-results used in here do not include the SRR step. Using the standard parametrization of BTK, the SRR outcome was less satisfying and of poorer quality than the reconstruction obtained via local neighborhood oriented Gaussian interpolation [Rousseau et al., 2006].

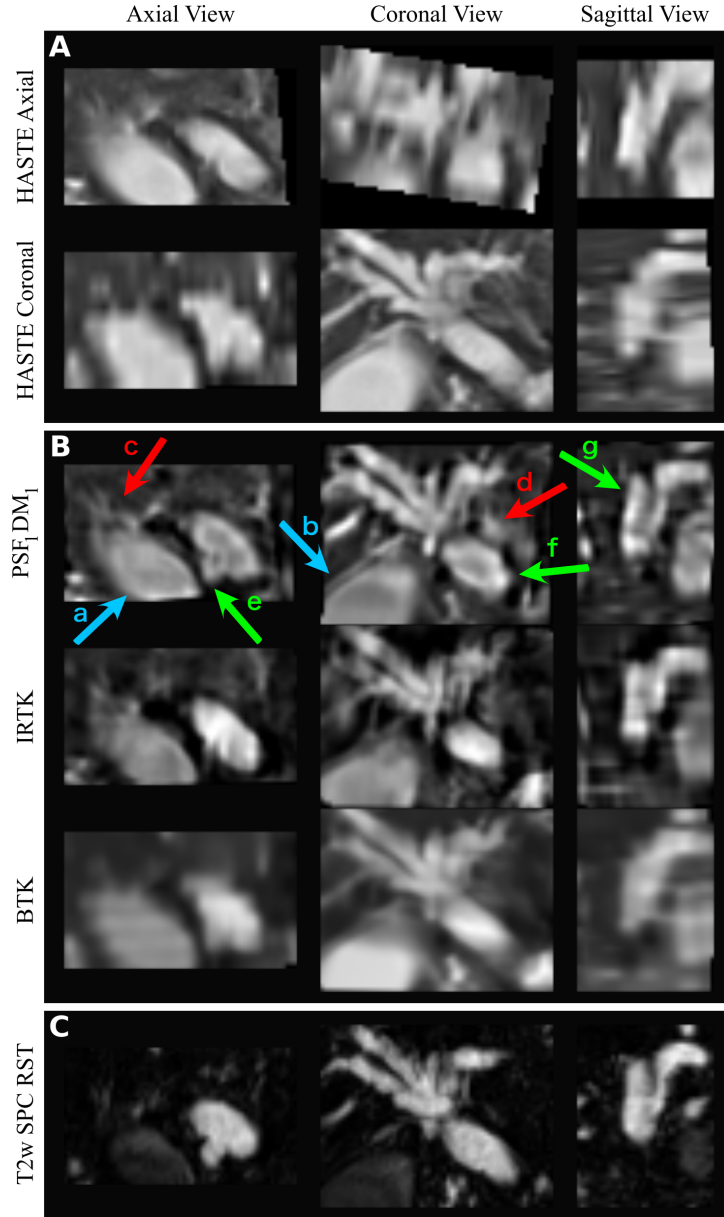


Figure 3.4.: Qualitative comparison between linearly resampled original HASTE data (A) and reconstructions obtained by BTK, IRTK and our proposed approach (B). Reconstructions are based on one axial and one coronal HASTE stack only. Several arrows on our reconstruction show examples of successfully preserved raw data information (blue *a* and *b*), introduction of artifacts (red *c* and *d*) and resolution improvement (green *e*, *f* and *g*) in direct comparison with the other reconstruction approaches. Artifacts are explained by similar intensities in the original data (*c*) in addition to the complex deformation occurred between axial and coronal stack acquisition (*d*). Resolution improvement was achieved by the combined usage of SR and the incorporated heavily T2-weighted volume information (C) as reference during motion correction.

3.3.2. Discussion

In this work, we present for the first time a single, consistent SRR framework which takes into consideration the PSF for both the motion correction and volumetric reconstruction steps. We put a particular focus on efficient implementation details like the matrix-free approach to efficiently compute the oriented Gaussian and adjoint oriented Gaussian operators for slice-to-volume registration and the PSF-aware volume-to-volume registration step. We test our framework by reconstructing upper abdominal MRI purely based on existing data available in current clinical MRCP studies. We propose a novel motion correction approach by using the existing heavily T2-weighted volume to guide the slice-to-volume registration to address the challenge of having only two orthogonal stacks with thick slices affected by deformable motion. Despite the high degree of undersampling, we achieve remarkable results which outperform current state-of-the-art techniques developed for fetal MRI, as shown in Figure 3.4. Further improvements in the current implementation include the incorporation of the oriented PSF for the gradient computation. This shortcoming could also describe the drop in accuracy for some slices observed in Figure 3.2. Overall, the obtained results are promising and may have the potential to avoid CT scans for further evaluation of this area. Existing limitations are assuming and only accounting for in-plane deformation and sparseness of available data. In Section 3.4, we expect to make further improvements by using more orthogonal stacks for higher anatomy sampling in combination with a more refined motion model. This will also allow increasing the field of view of the reconstruction to assess the entire biliary tree of clinical interest.

3.4. Super-Resolution for Upper Abdominal MRI: Acquisition and Post-Processing Protocol Optimisation using Brain MRI Control Data and Expert-Reader Validation

Our preliminary study [Ebner et al., 2017a] (Section 3.3) demonstrated the feasibility of upper abdominal MRI SRRs generated from only two standard MRCP protocol axial and coronal SST2W series using HT2W volumes as a reference-guide for in-plane deformable SVR/motion correction. However, anatomical clarity was lacking and a more robust registration/motion correction was needed. SRRs generated from a larger number of LR 2D source series are known to increase the reconstruction quality [Rousseau et al., 2010, Shilling et al., 2009] but acquisition of additional SST2W data comes at the expense of additional patient scanning time. Insight on the optimal orientation and number of input stacks for SRR is limited [Baker and Kanade, 2002, Lin and Shum, 2004, Candès and Fernandez-Granda, 2014], especially in the upper abdomen. Using HT2W volumes as a reference to guide registration is attractive but motion artefact arising from extended acquisition times and inconsistent breathing commonly degrades HT2W image quality. More rapidly acquired, similar to SST2W tissue contrast T2-weighted balanced fast-field echo (BFFE) sequences may offer a more consistent alternative for reference-guided registration, as may other more recently proposed non-reference guided SRR registration methods [Ebner et al., 2018c]. Finally, in order to objectively assess these factors, control studies using imaging free from significant variation in inter-subject motion artefact and from which a robust non-motion artefacted ground-truth/reference standard can be generated for SRR comparison are required.

In the pilot study presented in this section, we obtained healthy volunteer multiplanar SST2W stacks of the upper abdomen and the brain ('quasi-static' control data, to remove the effect of upper abdominal motion artefact), with the overall objectives of

- (a) optimising source image acquisition protocols by establishing the ideal num-

ber and orientation of SST2W series (so-called 'source data configuration') for MRCP SRR generation

- (b) optimising post-processing protocols by defining the best approach to registration/motion correction for SRR in the upper abdomen
- (c) validation of the overall potential of upper abdominal SRR, using expert-readers to compare pre-specified imaging features on the SRR with imaging obtained from standard SST2W MRCP protocols

3.4.1. Image Acquisition

Subjects and MRI Scanning

Local ethics committee approval was obtained and all participants provided informed written consent. Volunteers were recruited via advertisement within the University College London campus and were eligible if (a) they had no MRI contraindication, (b) were not taking any long-term medication (excluding the oral contraceptive pill) and (c) had no documented history of previous liver or gastrointestinal disease. The final cohort consisted of eight healthy volunteers (six male, mean age (28 ± 2) years, mean weight (72 ± 12) kg). Imaging was performed using a 3 T scanner (Ingenia, Philips Healthcare, Best, Netherlands) with a 16 channel body coil (SENSE XL Torso, Philips Healthcare, Best, Netherlands) used for abdominal imaging and a 15 channel head coil (dStream HeadSpine, Philips Healthcare, Best, Netherlands) used for brain imaging.

Image Acquisition Protocols

Upper Abdominal Imaging. Abdominal imaging was planned to ensure adequate coverage of the liver and biliary tree, with acquisition parameters listed in Table 3.2. Standard clinical axial and coronal SST2W series were acquired in expiratory breath-hold. The same acquisition parameters were used for additional expiratory breath-hold SST2W series planned in (a) the sagittal plane, (b) repeat axial, coronal and sagittal volumes shifted by half the slice thickness in the slice-select

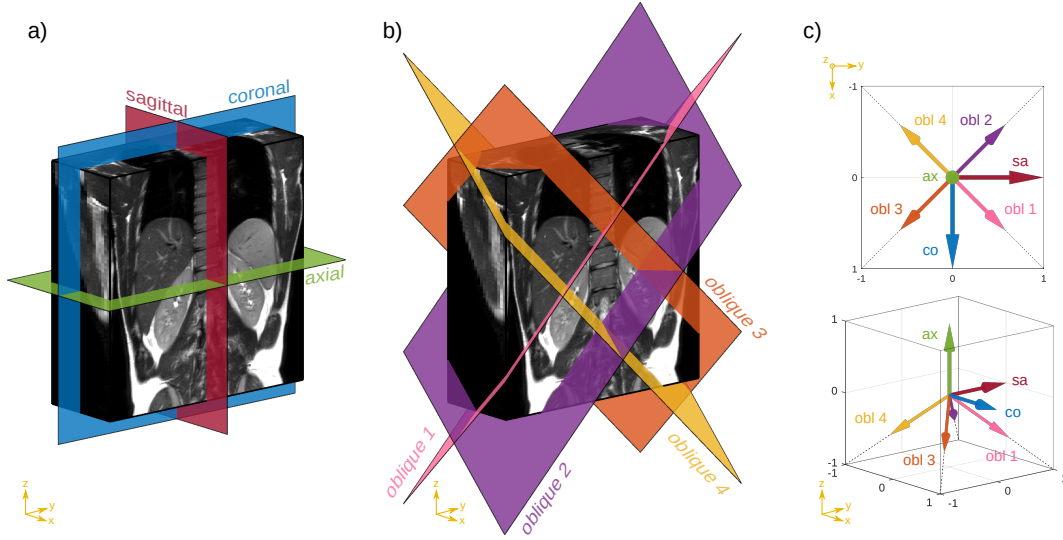


Figure 3.5.: Visualization of acquisition planes of SST2W images. Left and middle figures illustrate the imaging planes in the standard and the oblique orientations, respectively. The associated slice-select directions orthogonal to the respective acquisition planes are shown in the figure on the right. Example images are shown in Figure 3.6.

direction and (c) four additional oblique volumes where the slice-select dimensions were defined by the direction of a unit vector towards the lower four corners of a cube $[-1, 1]^3$ whose orientation is aligned with the standard anatomical directions (Figures 3.5 and 3.6).

For deformable reference-guided registration/motion correction, navigator-triggered free-breathing standard clinical HR HT2W volumes were acquired (Table 3.2). To investigate the potential of a less heavily T2-weighted volume for reference-guided registration that can be acquired more rapidly, we also used a thin slice 1.5 mm balanced fast field echo (BFFE) volume sequence to obtain high through-plane resolution coronal images (Table 3.2).

Quasi-static Control Brain Imaging. The same imaging protocol as presented for the abdomen was applied to the brain for seven out of the eight volunteers whereby identical imaging parameters were used to obtain image contrasts similar to the abdomen for the quasi-static control brain studies. All control data was planned to ensure adequate coverage of the brain whereby no imaging trigger was used. For

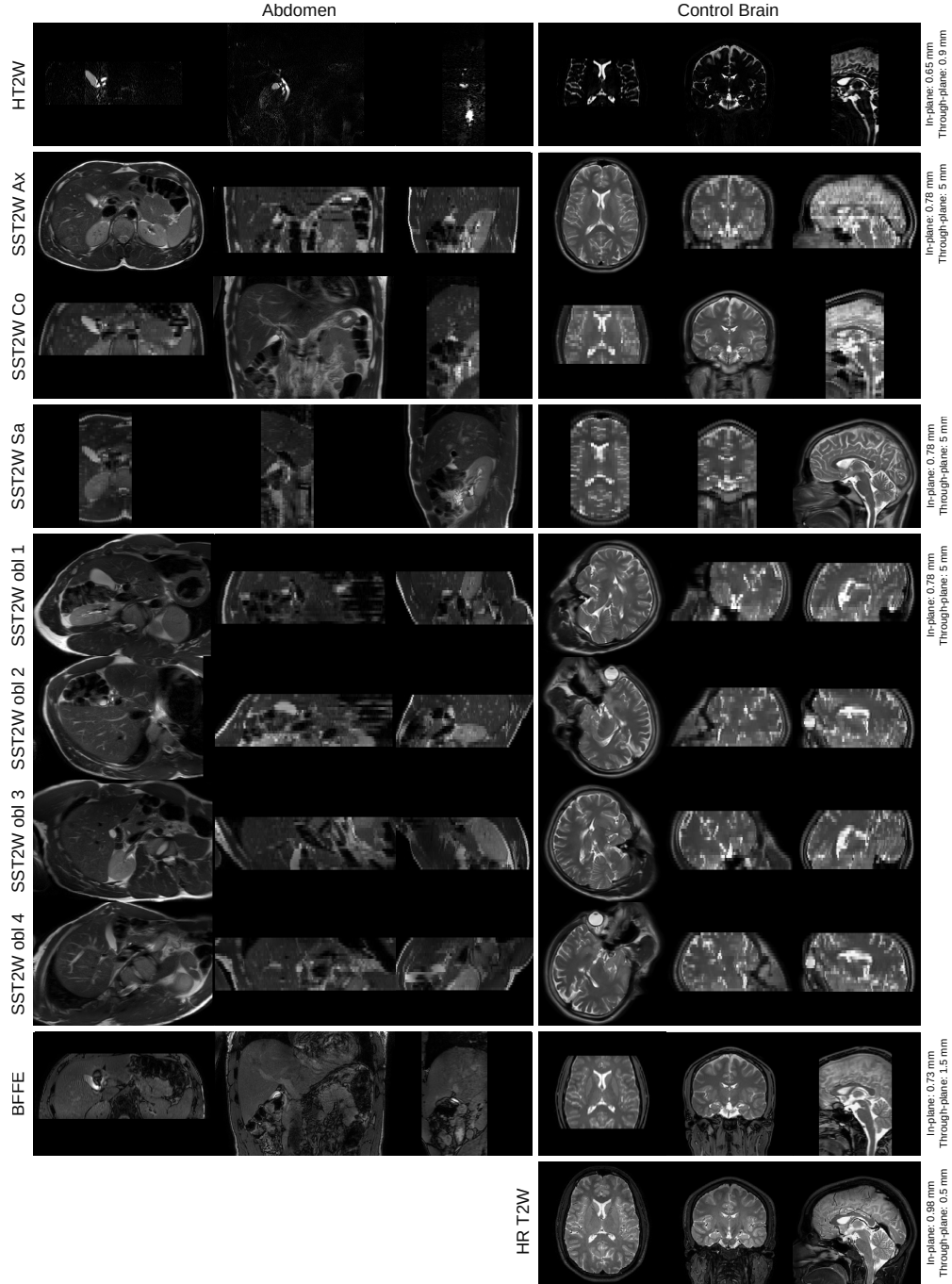


Figure 3.6.: Images obtained by extended MRCP protocol for abdomen and brain anatomies. The first three rows show the acquisitions that are available in standard clinical MRCP studies, i.e. an axial and a coronal SST2W images and an HT2W volume. Further acquisitions include SST2W images in sagittal and oblique orientations and a BFFE volume as an alternative candidate for the reference-guided motion correction framework. For validation purposes, a separate HR T2W volume was acquired for the brain.

Table 3.2.: Image acquisition protocol used in this volunteer study for both abdominal and control brain anatomies. For the abdominal imaging, the heavily T2-weighted (HT2W) volume is acquired as a gated acquisition triggered by a respiratory bellow. The single-shot T2-weighted (SST2W) stack and the balanced fast field echo (BFFE) acquisitions are acquired at separate expiratory breath-holds (BH). For the quasi-static control brain experiment, no imaging trigger are used and an additional HR T2W volume is acquired for ground-truth comparisons.

Example images associated with this protocol are shown in Figure 3.6.

Description	Abdomen and Control Brain			Control Brain only HR T2W volume
	HT2W volume	SST2W stack	BFFE volume	
Acquisition Type	3D	2D	3D	3D
Repetition Time [ms]	1120	1161	2.46	2500
Echo Time [ms]	662.00	80.00	1.23	252.83
Flip Angle [°]	90	90	15	90
Pixel Spacing [mm]	0.65×0.65	0.78×0.78	0.73×0.73	0.98×0.98
Slice Thickness [mm]	1.8	5	1.5	1
Slice Spacing [mm]	0.9	5	1.5	0.5
Number of Slices	90	20 – 25	83	360
Abdominal Imaging Trigger	Respiratory bellow	Expiratory BH	Expiratory BH	—
Abdominal Scan Duration	04:06.0	00:21.5 – 00:33.3	00:23.0	—

optimisation studies and ground-truth comparisons, an additional HR T2W volume was obtained (Table 3.2 and Figure 3.6).

3.4.2. Data, Evaluation Methodology and Results

Data and Data Preprocessing

The extended MRCP protocol as shown in Table 3.2 was applied for imaging both the abdomen and brain of eight healthy volunteers at University College London Hospital, London, UK. For one subject, no HR T2W volume was acquired of the brain which left seven subjects in a highly controlled, quasi-static environment for ground-truth comparisons. A ROI including the common bile duct, head of pancreas, porta hepatis and central liver was specified manually using masks generated on axial SST2W images by a radiologist with over ten years experience in abdominal imaging (PP). Quasi-static control ROIs for the brain tissue were defined automatically using the Brain Extraction Tool (BET) [Smith, 2002]. This region was also used for the quantitative ground-truth comparisons. SST2W images were preprocessed via an ITK bias field [Tustison et al., 2010] and a linear intensity correction step constrained by the provided manual (abdomen) and automatic (brain) masks.

Parametrisation of the Reconstruction Pipeline

For abdominal and quasi-static control brain data, the ROI mask was propagated to all the remaining SST2W series using nearest neighbour interpolation. Reconstruction pipelines were developed in PYTHON using ITK for the individual registration steps. Only data within the masked ROI was used for image processing and all slice registrations were constrained to the slice mask. For the deformable, reference-guided SRR framework, the in-plane deformation was performed using NIFTYREG³ software that is based on a fast free-deformation algorithm [Modat et al., 2010] using localized normalised cross-correlation (LNCC) as similarity measure. By applying the obtained in-plane deformation to each individual slice $\mathbf{y}_{s,i}$, the SRR problem (3.2) was solved using the transformed slices $\{\check{\mathbf{y}}_{s,i}\}_{s \in \mathbf{S}, i \in \mathbf{I}_s}$ in combination with the linear operators $\check{\mathbf{A}}_{s,i} = \mathbf{A}_{s,i}$, $s \in \mathbf{S}, i \in \mathbf{I}_s$, that carry the respective rigid slice motion correction estimates. To model the PSF of the image acquisition, we chose to approximate the SST2W sequence slice profile by a 3D Gaussian function defined by $\text{diag}\left(\frac{(1.2 s_1)^2}{8 \ln(2)}, \frac{(1.2 s_2)^2}{8 \ln(2)}, \frac{s_3^2}{8 \ln(2)}\right) \in \mathbb{R}^{3 \times 3}$ as variance-covariance matrix with s_1 , s_2 and s_3 representing the in- and through-plane spacings [Jiang et al., 2007, Kuklisova-Murgasova et al., 2012] in the slice-coordinate system, respectively. For the outlier-robust, rigid motion-correction and SRR framework, NIFTYMIC⁴ was used to solve (3.3) as described in [Ebner et al., 2018c]. Three iterations of two-step rigid SVR and outlier-robust SRR steps were performed with normalised cross-correlation (NCC) used to guide registrations. To create a first HR reference for the initial rigid SVRs, we used a discrete variant of Nadaraya-Watson kernel regression as an efficient scattered data approximation scheme for the volumetrically aligned SST2W image stacks [Vercauteren et al., 2006, Ebner et al., 2017a]. For outlier detection, Sim was set to NCC and β was selected empirically with values of 0.6, 0.65 and 0.7 per iteration to account for increasing accuracy in (3.3), respectively. There is broad consensus that SR in MRI can only reliably be achieved in through-plane and not in in-plane direction [Peled and Yeshurun, 2002, Plenge et al.,

³<https://github.com/KCL-BMEIS/niftyreg>

⁴<https://github.com/gift-surg/NiftyMIC>

2012, Scheffler, 2002, Van Reeth et al., 2012]. We, therefore, defined the isotropic reconstruction grid for the abdominal SRRs by the in-plane resolution of the stacks (0.78 mm). Given the SST2W slice thickness of 5 mm, our algorithm created an SR volume with approximately six times the resolution in the through-plane direction as the source SST2W images. For the brain, an isotropic reconstruction grid of 0.98 mm was used to approximately match the HR T2W volume resolution for the quantitative comparisons. The regularization parameter α was set to be 0.01 and 0.02 for the abdominal and quasi-static control SRRs, respectively. The different values are a result of the different reconstruction grid resolutions and were determined using a combination of L-curve studies [Hansen, 1992] and visual inspections.

Evaluation Methodology

Optimisation Control Studies for Brain MRI SRR. Six source data configurations for SRR generation were evaluated, using (i) axial and coronal ('a+c', two series); (ii) axial, coronal and sagittal ('a+c+s', three series); (iii) axial, coronal, sagittal and slice-select direction shifted axial, coronal and sagittal ('2a+2c+2s', six series); (iv) axial, coronal, sagittal and the first three oblique planes as shown in Figure 3.5 ('a+c+s+3obl', six series); (v) axial, coronal, sagittal and all four oblique planes ('a+c+s+4obl', seven series); and (vi) both axial, both coronal, both sagittal and all four oblique planes ('2a+2c+2s+4obl', ten series).

To evaluate the registration/motion-correction approaches, five SRRs were generated for each source data configuration using (i) no registration/motion correction (static SRR); (ii) reference-guided rigid registration using HT2W data (RG-HT2W); (iii) reference-guided rigid registration using BFFE data (RG-BFFE); (iv) reference-guided rigid registration using HR T2W data (RG-HRT2W); and (v) outlier-robust rejection rigid registration using only the SST2W source data used for each configuration (NiftyMIC). HR T2W brain imaging was used as ground-truth/reference standard imaging, with NCC used for both the similarity measure for each of the registration steps and the quantification of ground-truth similarity. Alternative similarity measures for registration (mutual information, MI; normalised mutual informa-

tion, NMI) and ground-truth similarity (structural similarity index measure [Wang et al., 2004], SSIM; NMI; peak-signal-to-noise ratio, PSNR) were also investigated and are presented in Figures B.2 and B.3 and Table B.1.

Analysis of RG-HRT2W based SRR results from the quasi-static control data is used to establish the optimal number and orientation of SST2W series for SRR while excluding the confounding factor of motion as much as possible.

Optimisation Studies for Upper Abdominal MRI SRR. The results obtained from the control brain studies were used to inform the abdominal imaging optimisation study methodology. Three source data configurations for SRR generation were evaluated: (i) axial and coronal ('a+c', two series), (ii) axial, coronal and sagittal ('a+c+s', three series), and (iii) axial, coronal, sagittal and the first three oblique planes ('a+c+s+3obl', six series). Four SRR approaches were evaluated for all previously utilised methods excluding the unavailable RG-HRT2W approach in this scenario.

In the absence of ground-truth/reference standard imaging, assessment was based on (i) numerical SRR self-consistency similarity measures, and (ii) subjective semi-quantitative analysis by two radiologists. Self-consistency was defined as the similarity between each slice $\mathbf{y}_{s,i}$ and its projected SRR counterpart $\mathbf{A}_{s,i}\mathbf{x}$ according to a similarity metric Sim, i.e. $\text{Sim}(\mathbf{y}_{s,i}, \mathbf{A}_{s,i}\mathbf{x})$, whereby NCC was used as Sim (SSIM, NMI and PSNR are additionally presented in Figure B.4 and Table B.3).

Subjective semi-quantitative evaluation was undertaken independently by two radiologists with over ten years experience in abdominal imaging (MC, PP) blinded to the SRR source data configuration or registration/motion-correction approach. The clarity of high signal intensity anatomical (biliary ductal) structures and presence of misregistration artefacts were scored as described in Section 3.3.1, and within-subject SRRs were ranked in order of preference. Where inter-reader discrepancies were noted, images were jointly re-evaluated and a consensus score was recorded after joint re-evaluation.

Because a reliable evaluation of 96 SRRs was not feasible, assessment was restricted to 24 SRRs at a time. The first experiment evaluated the three best-performing

image registration/motion-correction approaches determined from numerical self-consistency measures, in SRRs generated from the densest source data configuration (a+c+s+3obl, six series). The second experiment evaluated all three source data configurations for SRRs generated using the best-performing MRCP SRR approach determined from the first experiment.

Upper Abdominal MRI SRR Expert-Reader Validation Studies. Using the previously determined ideal number and orientation of SST2W series and the best approach for abdominal MRCP SRR, four radiologists (MC, PP, LF, ZA), three with subspeciality interest in hepato-pancreatico-biliary imaging and all with more than eight years experience in abdominal imaging, independently validated upper abdominal SRRs by direct comparison with standard protocol axial and coronal SST2W images, using a semi-quantitative scoring system. Both SRRs and standard SST2W images were scored for preservation of anatomical information at nine anatomical sites, focused predominantly on the assessment of peri-ductal and extra-ductal soft tissues. Readers also scored regions for the presence of artefacts, i.e. subjective but clinically apparent loss, addition or distortion of structures, introduced by SRR and recorded their subjective preference relative to standard SST2W images.

Statistical Analysis. Non-parametric statistical tests were used for the non-normally distributed NCC scores obtained from similarity measures between quasi-static control SRRs and ground-truth/reference standard imaging and for upper abdominal SRR self-consistency. This included Wilcoxon signed-rank tests for paired comparisons for the quasi-static control brain studies, and Kruskal-Wallis with post-hoc Dunn's tests for multiple comparisons for the abdominal optimisation studies. For the abdominal expert-reader experiments, non-parametric statistical tests were used for all reader-derived semi-quantitative scores, specifically Kruskal-Wallis tests with post-hoc Dunn's tests were used to determine differences between source data configurations or registration/motion-correction strategies for upper abdominal SRRs. For validation studies, Bland-Altman analysis of agreement for averaged clarity of anatomical information scores for SRR and standard SST2W imaging were compared

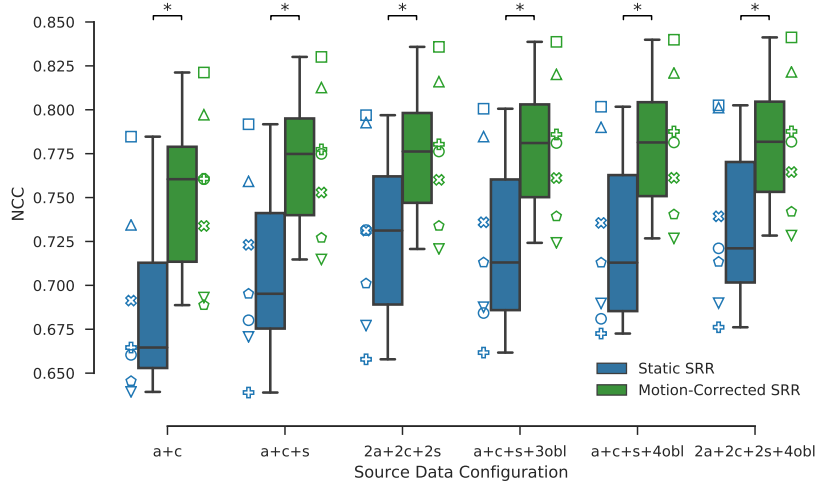


Figure 3.7.: Ground-truth (HR T2W) similarities for static and reference-guided SRR outcomes for the quasi-static control brain experiment whereby each of the seven subjects is assigned a different marker. The more input stacks are used the higher the similarity scores. Moreover, motion correction markedly improves the ground-truth similarities which was performed by rigidly registering each individual slice to the HR T2W volume using NCC as the similarity measure. A visual comparison for one subject is provided in Figure 3.8. Stars indicate statistical differences between the groups using a pairwise Wilcoxon signed-rank test ($p < 0.05$).

using the median difference as a bias measure and 2.5th and 97.5th percentiles as the 95% Limits-of-Agreement (LoA). Wilcoxon signed-rank tests were used to test for differences of anatomical clarity between SRRs and standard clinical axial and coronal SST2W data ('Ax&Co'), subjective preference and the presence of visible artefacts. The threshold of statistical significance for all tests was defined as $p < 0.05$.

Results

Optimisation Control Studies for Brain MRI SRR. A total of 210 quasi-static control brain SRRs were generated (30 SRRs for each of the seven volunteers). The box-plot in Figure 3.7 illustrates the impact of motion correction and source data configuration on the NCC, SSIM, NMI and PSNR scores. It shows that adding more than six input stacks (a+c+s+3obl) leads to only little numerical improvement in reconstruction quality for MRCP SST2W data. Additional comparisons including

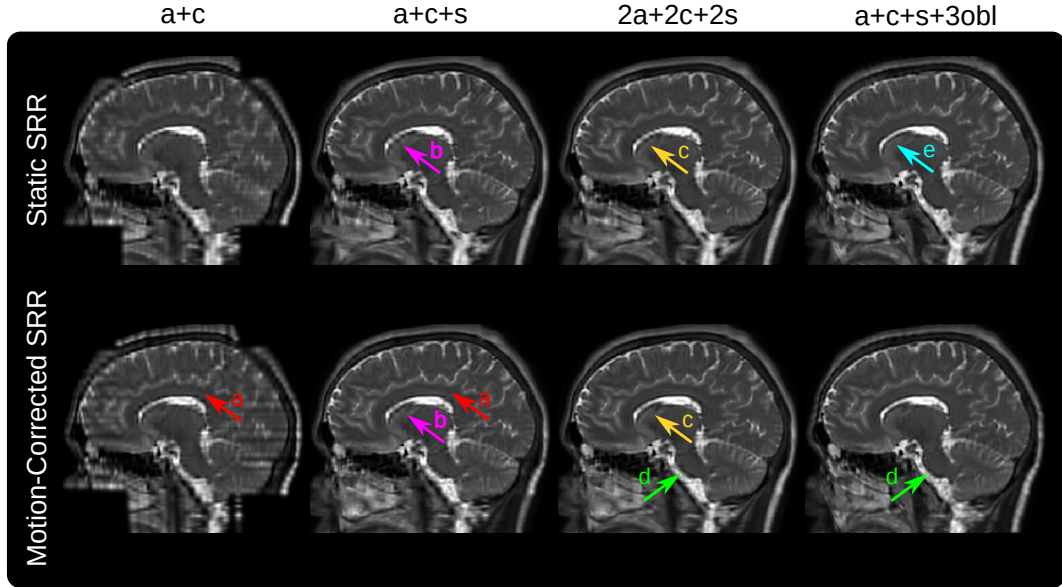


Figure 3.8.: Qualitative comparison of the static and reference-guided SRR outcome of one subject for various input data scenarios in the sagittal view (additional axial and coronal view comparisons are shown in Figure B.1). It illustrates the impact of the number of input stacks and how multiple orientations can improve PVE recovery. In particular, SRR (a+c+s+3obl) shows visually higher anatomical accuracy than SRR (2a+2c+2s) despite the same number of six input stacks used for the SRR. The red arrows (a) underline that the SRR based on only two stacks (a+c) as currently available for clinical MRCP study protocols produces a very poor SRR quality which is especially noticeable in the sagittal view. The magenta arrows (b) illustrate that for three input stacks (a+c+s) the corpus callosum can only be reconstructed with limited geometrical integrity. Motion-correction helps to recover it more clearly by adding three additional stacks (2a+2c+2s) as indicated by arrows (c). The green arrows (d) show the improved visual clarity at the medulla due to better PVE correction by using oblique data. Additional oversampling for high input stack numbers leads to higher PSNR. This may also result in clear tissue boundaries even in case of insufficient motion correction for the static SRR as indicated by the cyan arrow (e).

Table 3.3.: Ground-truth (HR T2W) NCC-similarities of obtained quasi-static control brain SRRs for an increasing number of input stacks for different motion correction (MC) strategies summarised for all seven subjects. The rows are sorted in a descending order according to the SRR outcome for 'a+c+s+3obl'.

MC Strategy	a+c	a+c+s	2a+2c+2s	a+c+s+3obl
RG-HRT2W	0.751 ± 0.046	0.770 ± 0.039	0.775 ± 0.038	0.779 ± 0.038
RG-BFFE	0.735 ± 0.047	0.754 ± 0.039	0.759 ± 0.038	0.764 ± 0.038
NiftyMIC	0.724 ± 0.052	0.748 ± 0.043	0.751 ± 0.041	0.758 ± 0.040
RG-HT2W	0.708 ± 0.042	0.734 ± 0.037	0.739 ± 0.037	0.750 ± 0.037
Static SRR	0.689 ± 0.049	0.708 ± 0.049	0.727 ± 0.050	0.724 ± 0.049

all motion-correction strategies are provided in Figure B.3.

A visual comparison in Figure 3.8 illustrates how different source data configurations affect the SRR results. In particular, using two input stacks (a+c) leads to inferior outcomes which is especially noticeable in the sagittal plane. Three input stacks (a+c+s) yield a substantial visual improvement that is further visible by adding three more stacks as shown for both 2a+2c+2s and a+c+s+3obl SRR outcomes. Although relatively subtle in this comparison, using oblique orientations (a+c+s+3obl) instead of same-plane acquisitions (2a+2c+2s) can lead to more accurate SRRs depending on the curvature of tissue structures. This is indicated at the medulla which appears more blurred for the 2a+2c+2s outcome. Adding more stacks shows little visual improvement. However, the additional oversampling leads to higher PSNR and may result in clearer tissue boundaries (Figure 3.8).

A quantitative comparison of the five registration/motion-correction approaches with respect to different source data configurations is shown in Table 3.3. Only a subset of all performed comparisons is provided for simplicity (further comparisons are available in Table B.1 which also includes using the axial SST2W stack (SST2Wax) as another possible choice as reference image for the reference-guided SRR approach). The RG-HRT2W outcome represents an approximation of the upper bound for the theoretically achievable MRCP SRR quality. Of the remaining four SRR approaches, using the BFFE volume as a reference performs second best. NiftyMIC, which does not rely on any external reference, performs between RG-BFFE and RG-HT2W. NiftyMIC demonstrated negligible outlier rejections (average 0.05 ± 0.21 slices) during SRR. In terms of source data configurations, numerical outcomes confirm the importance of multiplanar image input for high SRR quality. In particular, oblique planes (a+c+s+3obl) are preferable over multiple standard planes (2a+2c+2s).

Optimisation Studies for Upper Abdominal MRI SRR. Based on the findings from the control quasi-static brain MR data, we tested three source data configurations (a+c, a+c+s, a+c+s+3obl) using all four registration/motion-correction methods available in the abdomen (RG-BFFE, RG-HT2W, NiftyMIC and static SRR). A total of 96 abdominal SRRs were generated (12 SRRs for each of the

Table 3.4.: Clinical evaluation by two radiologists for abdomen averaged over all eight subjects for two separate experiments: A first experiment based on six input stacks (a+c+s+3obl), and a second experiment involving three source data combinations (a+c, a+c+s, a+c+s+3obl) for the best performing method in the first experiment (NiftyMIC). Clarity of anatomical structures score indicates how well common bile duct (CBD), and left and right hepatic ducts (LHD and RHD) are visualized in each image with ratings 0 (structure not seen), 1 (poor depiction), 2 (suboptimal visualization; image not adequate for diagnostic purposes), 3 (clear visualization of structure but reduced tissue contrast; image-based diagnosis feasible) and 4 (excellent depiction; optimal for diagnostic purposes). Visible motion score rates the amount of visible non-corrected motion from score 0 (complete motion) to 3 (no motion). Radiologists' preference ranks the subjectively preferred reconstructions from 1 to 3 (least to most preferred). Stars in the last three columns indicate that the respective score is statistically significantly different from the respective other two groups based on Kruskal-Wallis and post-hoc Dunn tests ($p < 0.05$).

		Clarity of Anat. Structures			Total Score Anat. Clarity	Visible Motion	Radiologists' Preference
		CBD	LHD	RHD			
Exp. 1: a+c+s+3obl	Static SRR	1.25±0.46	1.25±0.46	1.25±0.46	3.75±1.16	1.25±0.46	1.62±0.52
	RG-HT2W	1.50±0.76	1.62±0.52	1.62±0.52	4.75±1.58	1.25±0.71	1.38±0.52
	NiftyMIC	2.88±0.35	2.62±0.52	2.62±0.52	8.12±1.25*	2.25±0.71*	3.00±0.00*
Exp. 2: NiftyMIC	a+c	1.38±0.52	1.12±0.35	1.12±0.35	3.62±0.74	1.00±0.00	1.12±0.35
	a+c+s	1.75±0.46	1.50±0.53	1.50±0.53	4.75±0.89	1.38±0.52	1.88±0.35
	a+c+s+3obl	2.62±0.52	2.50±0.76	2.25±0.71	7.38±1.41*	2.12±0.35	3.00±0.00*

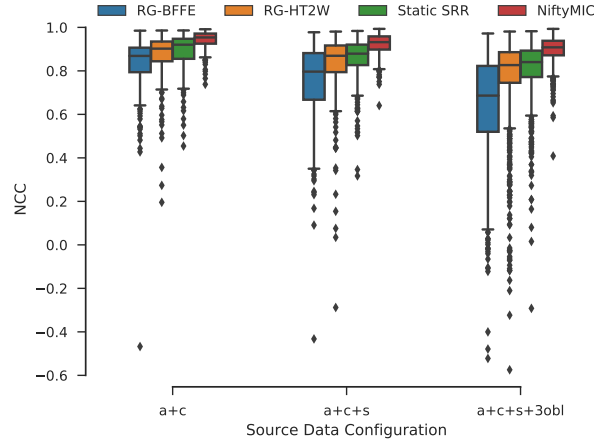


Figure 3.9.: Self-consistency evaluation given by projected NCC-similarities for all slices of obtained abdominal SRRs for an increasing number of input stacks for different motion-correction strategies summarised for all eight subjects. All self-consistency outcomes between SRR approaches, except for RG-HT2W vs Static SRR for 'a+c+s', are significantly different within each source data configuration based on Kruskal-Wallis with post-hoc Dunn tests ($p < 0.05$).

eight volunteers). The scores in Figure 3.9 indicate highest self-consistency for the NiftyMIC SRR outcomes across source data configurations for NCC followed by Static SRR (more comparisons including SSIM, NMI and PSNR as self-consistency similarity measures in addition to a comparison against a rigid-only variant for the reference-guided approach, i.e. no in-plane deformation step, are provided in Table B.3). Lower outcomes for the reference-guided approaches indicate the existence of slice misregistrations with RG-HT2W producing consistently better results. For NiftyMIC, the statistics for the outlier-robust framework were ($\#$ of rejected slices / $\#$ of total slices) $0.14 \pm 0.35 / 30.43 \pm 4.40$, $0.71 \pm 0.88 / 50.29 \pm 4.33$, and $5.00 \pm 2.14 / 108.57 \pm 4.87$ for the a+c, a+c+s and a+c+s+3obl input data scenarios, respectively, indicating a moderate increase of slice rejections for increasing input data.

Figure 3.10 shows that SRRs based on fewer input stacks represent anatomically less plausible reconstructions. Both RG-HT2W and NiftyMIC show improved anatomical clarity over the static SRR approach. This is especially the case for the SRRs based on six input stacks. However, RG-HT2W becomes less accurate in areas with poor HT2W image contrast.

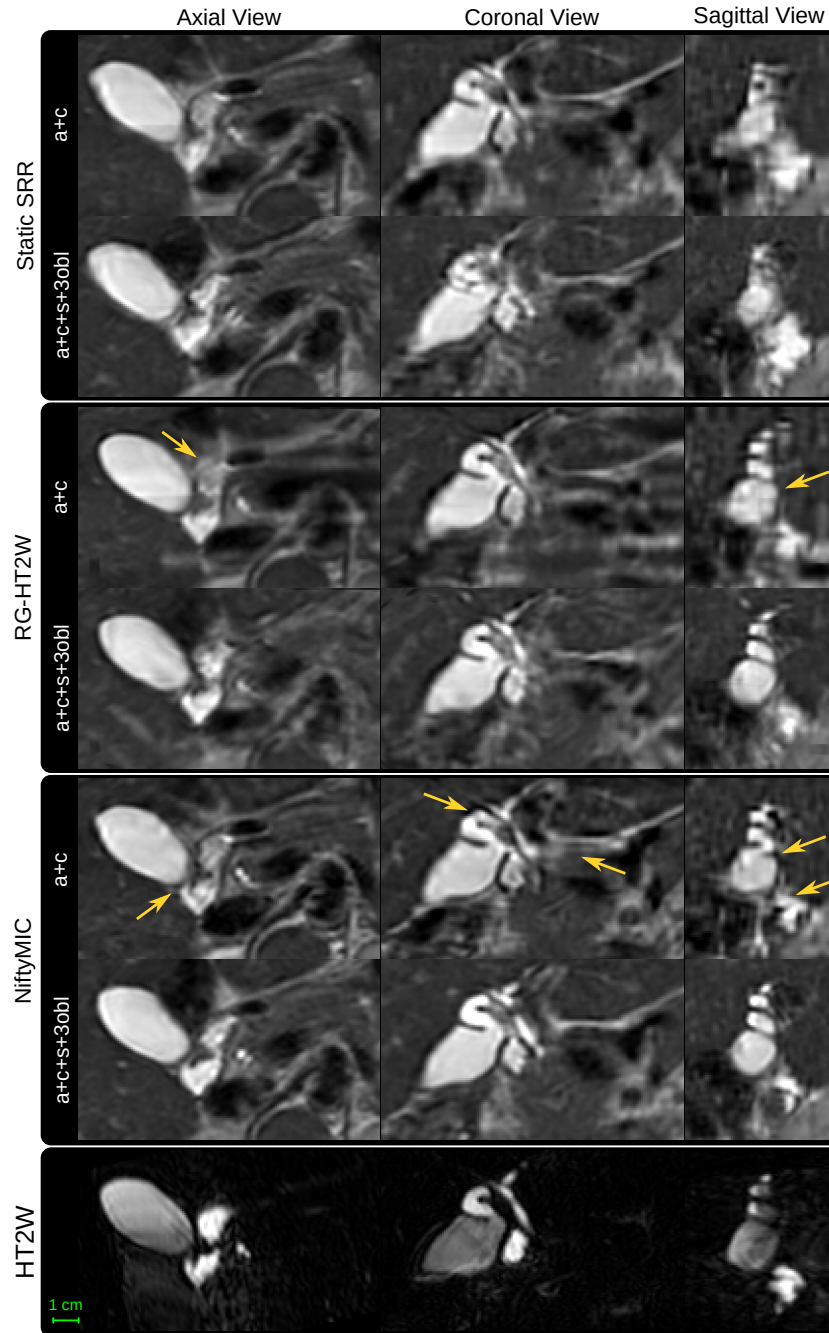


Figure 3.10.: Qualitative comparison between the SRR approaches using either two or six input stacks. Both motion-correction frameworks, i.e. the HT2W-guided one and NiftyMIC, achieve SRRs with visually improved anatomical plausibility. However, in areas where MRCP lacks contrast, NiftyMIC tends to produce superior results. Moreover, using six input stacks can lead to better SRR outcomes in case of adequate motion correction which is especially visible in the sagittal view.

Examples for such visual improvements are indicated by arrows.

Table 3.4 summarises the two radiologists’ qualitative evaluation. Superiority of NiftyMIC over RG-HT2W and Static SRR in terms of anatomical clarity, amount of visible motion and the radiologists’ preference was statistically significant. By selecting NiftyMIC as the best-performing MRCP SRR approach, a second experiment demonstrated the significantly better reconstruction quality achieved by using the a+c+s+3obl source data. In particular, it was selected as the radiologists’ preference without exception. Two independently conducted experiments with, at least, days delay, show little variability in the radiologists’ assessment of the NiftyMIC a+c+s+3obl outcome, Table 3.4 rows 3 and 6.

Upper Abdominal MRI SRR Expert-Reader Validation Studies. NiftyMIC with a+c+s+3obl input stacks was chosen as best-performing method for subsequent validation studies. Bland-Altman analysis of agreement between SRR and standard axial and coronal SST2W data in Figure 3.11 confirms a systematically better outcome in clarity of anatomical structures presented on the SRRs. Assessment of the individual anatomical regions shows a statistically better SRR performance for two of the nine assessed anatomical features (cystic duct, and hepatic artery and central branches). Expert-reader subjective preference scores demonstrate statistically significant preferences for vascular structures, the cystic duct and first generation intrahepatic ducts. Pancreatic duct, head-of-pancreas parenchyma and ampulla were preferred on standard imaging. Importantly, artefact scores were generally well above 1 which underlines that the SRRs present minimal or no new artefact in comparison with the original data. On average, no structure was scored as having less artefact than the original data.

PSF-Aware Slice-to-Volume Registration. For the quasi-static control brain data, additional experiments were performed to investigate the potential benefit of using PSF-aware registration. The box-plots in Figure 3.12 illustrate the differences in SRR quality obtained with and without PSF-awareness for the rigid slice-to-volume motion correction based on RG-HRT2W-NCC, i.e using NCC as similarity measure for RG-HRT2W. The comparison in Table 3.5 summarizes the impact of

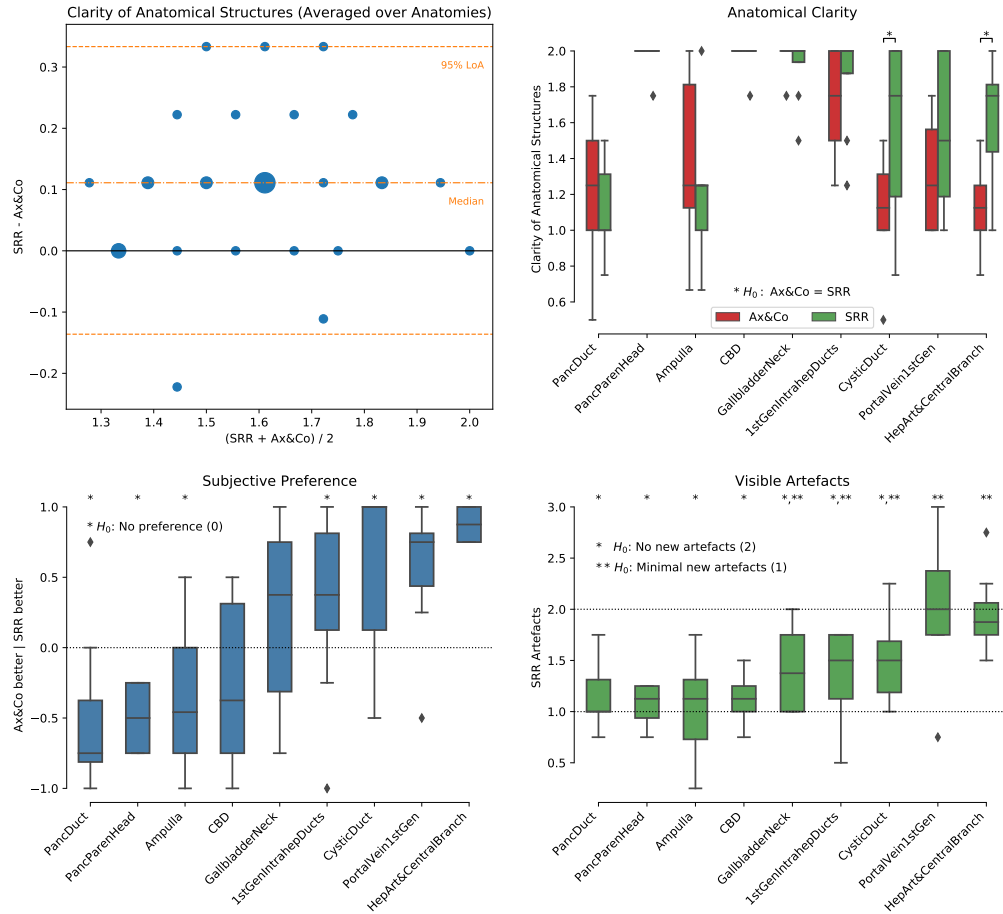


Figure 3.11.: Clinical evaluation by four radiologists for the abdomen of the third experiment for all eight subjects. Top panels: Clinical interpretability scores were 0 (structure not identified), 1 (structure poorly visualized) and 2 (structure clearly visualized). Bottom left: Subjective impression score ranks how frequently the SRR (NiftyMIC a+c+s+3obl) was considered subjectively of worse, same or better quality than the original axial and coronal SST2W data for interpretation. Bottom right: Artefacts measure to what extent the SRR presented additional artefacts with respect to the original axial and coronal SST2W data with scores 0 (lots of new artefacts), 1 (minimal new artefacts), 2 (no new artefact) and 3 (less artefact than original). Stars are shown to illustrate significant outcomes by rejecting the respective H_0 hypothesis based on a Wilcoxon signed-rank test ($p < 0.05$).

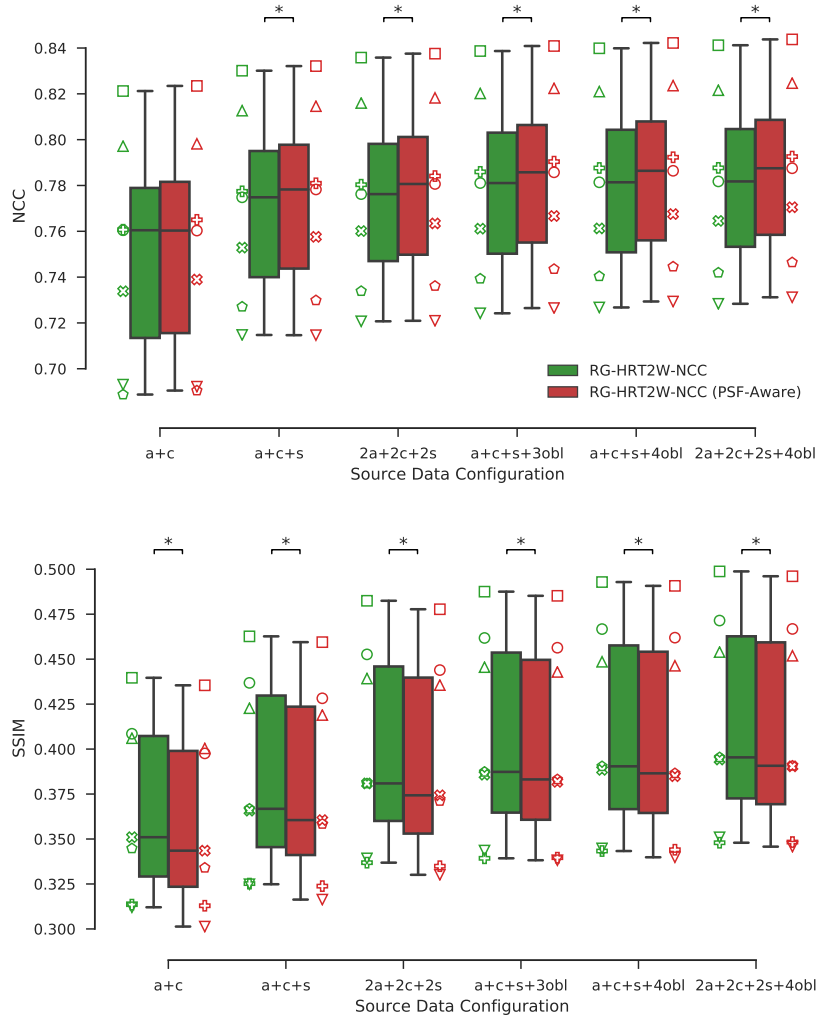


Figure 3.12.: Evaluation of ground-truth (HR T2W) similarities for PSF-aware registration. Ground-truth (HR T2W) similarities for standard and PSF-aware reference-guided SRR outcomes for the quasi-static brain experiment involving seven subjects. Stars indicate statistical differences between the groups using a pairwise Wilcoxon signed-rank test ($p < 0.05$). The PSF-aware registration leads to higher NCC-similarities across input source configurations. However, this trend cannot be confirmed in terms of SSIM which is further underlined in Table 3.5.

using PSF-awareness during registration for additional motion correction strategies whereby only the SRR using the source data configuration 'a+c+s+3obl' was considered for its potential relevance in MRCP SRR. However, whereas a PSF-awareness in the registration tends to achieve improved ground-truth similarities in terms of measured NCC and PSNR scores, this trend cannot be confirmed systematically.

Table 3.5.: Evaluation of ground-truth similarities for PSF-aware registration.

Ground-truth (HR T2W) similarities of the 'a+c+s+3obl'-SRR outcome of obtained quasi-static control brain SRRs with and without PSF-awareness for different reference-guided motion correction (MC) strategies summarised for all seven subjects. Using the short-hand

“RG-Reference-SimilarityMeasureForRegistration”, the settings for the reference-guided SRR approach are shown, where, e.g., RG-BFFE-NMI refers to the use of BFFE as reference volume for guidance and NMI as similarity measure for registration. The rows are sorted in a descending order according to the PSF-aware-based result as measured in NCC (second column). Stars indicate statistical differences between the groups using a pairwise Wilcoxon signed-rank test ($p < 0.05$).

MC Strategy	NCC		SSIM		PSNR	
	Standard	PSF-Aware	Standard	PSF-Aware	Standard	PSF-Aware
RG-HRT2W-NCC	0.78±0.04	0.78±0.04*	0.41±0.06*	0.40±0.06	14.35±3.29	14.43±3.29*
RG-HRT2W-NMI	0.77±0.04	0.77±0.04*	0.41±0.06	0.41±0.06*	14.25±3.27	14.26±3.29
RG-BFFE-NCC	0.76±0.04	0.77±0.04*	0.40±0.06	0.40±0.06	14.38±3.29	14.41±3.29*
RG-BFFE-NMI	0.76±0.04	0.77±0.04*	0.40±0.06	0.41±0.06*	14.33±3.30	14.37±3.31*
RG-HRT2W-MI	0.78±0.04*	0.76±0.05	0.39±0.06*	0.34±0.06	14.47±3.27	14.67±3.14
RG-HT2W-NCC	0.75±0.04*	0.74±0.04	0.39±0.05*	0.39±0.05	14.15±3.11*	14.05±3.08
RG-SST2Wax-NCC	0.74±0.04	0.72±0.04	0.38±0.05	0.37±0.04	14.13±3.21	13.50±3.08
RG-BFFE-MI	0.48±0.20	0.62±0.15*	0.14±0.16	0.23±0.15*	14.41±2.71	14.72±2.93
RG-SST2Wax-MI	0.41±0.06	0.50±0.04*	0.10±0.02	0.14±0.03*	13.85±2.13	13.88±2.46
RG-HT2W-MI	0.29±0.10	0.41±0.08*	0.07±0.03	0.12±0.03*	13.53±2.02	13.89±2.09*

Table 3.6.: Evaluation of slice-similarities for PSF-aware registration.

Slice-similarities $\text{Sim}(\mathbf{y}_{s,i}, \mathbf{A}_{s,i}\mathbf{x}_{\text{ref}})$ between slices $\mathbf{y}_{s,i}$ and the projection $\mathbf{A}_{s,i}\mathbf{x}_{\text{ref}}$ from the reference volume \mathbf{x}_{ref} that was used by the respective reference-guided motion correction (MC) strategy with and without PSF-awareness summarised for all slices of the seven subjects. Using the short-hand

“RG-Reference-SimilarityMeasureForRegistration”, the settings for the reference-guided SRR approach are shown, where, e.g., RG-BFFE-NMI refers to the use of BFFE as reference volume for guidance and NMI as similarity measure for registration. The rows are sorted according to Table 3.5. Stars indicate statistical differences between the groups using a pairwise Wilcoxon signed-rank test ($p < 0.05$).

MC Strategy	NCC		SSIM		PSNR	
	Standard	PSF-Aware	Standard	PSF-Aware	Standard	PSF-Aware
RG-HRT2W-NCC	0.79±0.08	0.79±0.08*	0.46±0.07*	0.45±0.07	13.34±4.17	13.39±4.17*
RG-HRT2W-NMI	0.78±0.09	0.78±0.09*	0.46±0.07	0.47±0.07*	13.23±4.21	13.29±4.18*
RG-BFFE-NCC	0.87±0.04	0.88±0.04*	0.40±0.06*	0.40±0.06	11.26±3.32	11.25±3.30
RG-BFFE-NMI	0.87±0.05	0.87±0.04*	0.40±0.06	0.40±0.06*	11.26±3.35	11.28±3.32*
RG-HRT2W-MI	0.76±0.12*	0.71±0.17	0.43±0.09*	0.37±0.13	13.25±4.10*	13.04±4.05
RG-HT2W-NCC	0.84±0.09	0.85±0.09*	0.18±0.06*	0.18±0.06	7.66±2.54*	7.54±2.61
RG-SST2Wax-NCC	0.80±0.12*	0.79±0.12	0.44±0.23*	0.43±0.22	23.71±5.09*	23.60±4.89
RG-BFFE-MI	0.46±0.33	0.58±0.33*	0.16±0.17	0.22±0.17*	10.64±3.25	10.86±3.27*
RG-SST2Wax-MI	0.37±0.28	0.46±0.28*	0.15±0.18	0.19±0.20*	18.22±2.91	18.98±3.26*
RG-HT2W-MI	0.19±0.29	0.30±0.37*	0.03±0.08	0.06±0.09*	6.93±2.53	7.12±2.49*

A similarly inconclusive outcome can be observed by assessing the slice-similarities

$\text{Sim}(\mathbf{y}_{s,i}, \mathbf{A}_{s,i}\mathbf{x}_{\text{ref}})$ between slices $\mathbf{y}_{s,i}$ and the projection $\mathbf{A}_{s,i}\mathbf{x}_{\text{ref}}$ from the respective

reference volume \mathbf{x}_{ref} used to guide the registration as shown in Table 3.6. PSF-aware registration generally achieves higher similarities for the slice motion correction estimates in terms of NCC, but not necessarily in terms of SSIM or PSNR.

Abdominal MRI SRR using Total Variation Regularization. To investigate a potential improvement of the volumetric reconstruction quality by using a different regularizer other than the proposed first-order Tikhonov regularization (TK1), we additionally performed comparisons against isotropic total variation (TV) regularization. Thus, we compared the obtained SR reconstructions

$$\mathbf{x}^* := \arg \min_{\mathbf{x} \geq 0} \left(\sum_{s \in \mathbf{S}} \sum_{i \in \mathbf{I}_s} \frac{1}{2} \|\mathbf{y}_{s,i} - \mathbf{A}_{s,i} \mathbf{x}\|_{\ell^2}^2 + \alpha \Psi(\mathbf{x}) \right) \in \mathbb{R}^N \quad (3.4)$$

using both the TK1 and TV regularizers defined as

$$\Psi(\mathbf{x}) = \text{TK}_1(\mathbf{x}) := \frac{1}{2} \|\nabla \mathbf{x}\|_{\ell^2}^2 := \frac{1}{2} \sum_{k=1}^N (\partial_x \mathbf{x}(k))^2 + (\partial_y \mathbf{x}(k))^2 + (\partial_z \mathbf{x}(k))^2 \quad (3.5)$$

and

$$\Psi(\mathbf{x}) = \text{TV}_{\text{iso}}(\mathbf{x}) := \|\nabla \mathbf{x}\|_{\ell^1} := \sum_{k=1}^N \sqrt{(\partial_x \mathbf{x}(k))^2 + (\partial_y \mathbf{x}(k))^2 + (\partial_z \mathbf{x}(k))^2}, \quad (3.6)$$

respectively. However, while the TK1 problem can be solved efficiently using a linear least-squares formulation the TV formulation requires a more complex framework that can deal with the associated non-smooth (but still convex) optimisation problem. For the implementation of the isotropic TV solver we used a primal-dual (PD) algorithm presented in [Chambolle and Pock, 2011] known for its suitability and fast convergence in imaging problems. Our PD solver implementation is publicly available in the NSoL⁵ package which is integrated into NIFTYMIC.

To allow for a direct comparison of the TV and TK1 regularizer outcomes, their respective SRR problem (3.4) was solved after performed motion correction of the respective MRCP SRR frameworks as presented in the Sections 3.2.1 and 3.2.2 (SRR

⁵<https://github.com/gift-surg/NSoL>

using Reference-Guided Multimodal Deformable Motion Correction and Outlier-robust SRR using Monomodal Rigid Motion Correction). Therefore, for the non-iterative reference-guided approach, (3.4) was solved after finishing the individual slice registrations. For the iterative two-step registration-reconstruction framework NiftyMIC, the respective regularizer is applied before the final SRR reconstruction step. This process allows a direct comparison of the TV and TK1 regularizers on the obtained SRR outcome without the confounding factor of different motion correction estimates. Additionally, this helps to keep the computational times low for NiftyMIC as the TV problem only needs to be solved once at the end of the two-step iterations.

Similar to the parametrisation of the TK1-based reconstruction pipeline described in the Section 3.4.2 (Data and Data Preprocessing), parameter studies were performed to determine suitable TV-regularization parameters α and the number of required PD iterations to achieve convergence. Based on additional visual inspection, TV-regularization parameters of $\alpha \in \{0.0001, 0.0005, 0.0009\}$ were chosen for the comparisons in here. Considering the input source data configurations of 'a+c', 'a+c+s' and 'a+c+s+3obl', it was found that 15 PD iterations are sufficient to achieve convergence for the chosen regularization parameters.

Table 3.7 provides a direct comparison of the obtained ground-truth (HR T2W) similarities for the quasi-static control brain experiment using TK1 and TV regularization for the SRR after RG-HRT2W-NCC-based motion correction. A visual summary of the outcomes is also provided in Figure 3.13 comparing TK1 and TV using $\alpha = 0.0005$. The comparisons show that TV does not lead to an improvement of the SRR as quantified by the similarity measures of NCC, SSIM and NMI. However, TV shows slightly increased PSNR compared to TK1. A qualitative comparison in Figure 3.14 represents an extension to Figure 3.10 and shows that TV produces visually similar reconstructions compared to TK1 for low regularization parameters α . Larger α values for TV can lead to slightly sharper contours but a delicate balance needs to be found in order to avoid the introduction of staircase artefacts typical for TV which may well suppress clinically relevant structural information.

Finally, typical computational times to reconstruct a HR volume around the biliary

Table 3.7.: Ground-truth (HR T2W) similarities of obtained quasi-static control brain SRRs using first-order Tikhonov (TK1) and isotropic Total Variation (TV) regularization SRR outcomes for an increasing number of input stacks for all seven subjects. The respective regularization was applied in the final reconstruction step using RG-HRT2W-NCC as the motion-correction strategy.

(a) NCC						
Regularizer	a+c	a+c+s	2a+2c+2s	a+c+s+3obl	a+c+s+4obl	2a+2c+2s+4obl
TK1	0.751 ± 0.046	0.770 ± 0.039	0.775 ± 0.038	0.779 ± 0.038	0.780 ± 0.038	0.781 ± 0.038
TV ($\alpha=9e-4$)	0.730 ± 0.043	0.750 ± 0.037	0.756 ± 0.036	0.761 ± 0.036	0.763 ± 0.036	0.764 ± 0.035
TV ($\alpha=5e-4$)	0.727 ± 0.043	0.748 ± 0.037	0.755 ± 0.036	0.759 ± 0.036	0.761 ± 0.036	0.763 ± 0.035
TV ($\alpha=1e-4$)	0.723 ± 0.043	0.745 ± 0.037	0.753 ± 0.036	0.757 ± 0.036	0.760 ± 0.036	0.762 ± 0.035

(b) SSIM						
Regularizer	a+c	a+c+s	2a+2c+2s	a+c+s+3obl	a+c+s+4obl	2a+2c+2s+4obl
TK1	0.368 ± 0.047	0.386 ± 0.051	0.402 ± 0.053	0.407 ± 0.054	0.411 ± 0.055	0.416 ± 0.055
TV ($\alpha=9e-4$)	0.325 ± 0.039	0.348 ± 0.041	0.360 ± 0.044	0.363 ± 0.044	0.367 ± 0.045	0.371 ± 0.045
TV ($\alpha=5e-4$)	0.320 ± 0.040	0.342 ± 0.043	0.356 ± 0.045	0.359 ± 0.045	0.363 ± 0.046	0.368 ± 0.046
TV ($\alpha=1e-4$)	0.309 ± 0.041	0.330 ± 0.044	0.349 ± 0.046	0.351 ± 0.046	0.355 ± 0.046	0.363 ± 0.047

(c) NMI						
Regularizer	a+c	a+c+s	2a+2c+2s	a+c+s+3obl	a+c+s+4obl	2a+2c+2s+4obl
TK1	1.090 ± 0.014	1.103 ± 0.015	1.107 ± 0.016	1.110 ± 0.016	1.111 ± 0.016	1.112 ± 0.016
TV ($\alpha=9e-4$)	1.086 ± 0.012	1.097 ± 0.013	1.100 ± 0.014	1.103 ± 0.014	1.104 ± 0.014	1.104 ± 0.014
TV ($\alpha=5e-4$)	1.083 ± 0.012	1.095 ± 0.013	1.098 ± 0.014	1.100 ± 0.014	1.101 ± 0.014	1.102 ± 0.014
TV ($\alpha=1e-4$)	1.080 ± 0.011	1.091 ± 0.012	1.095 ± 0.013	1.097 ± 0.013	1.099 ± 0.013	1.100 ± 0.014

(d) PSNR						
Regularizer	a+c	a+c+s	2a+2c+2s	a+c+s+3obl	a+c+s+4obl	2a+2c+2s+4obl
TK1	14.275 ± 3.048	14.230 ± 3.027	14.276 ± 3.011	14.349 ± 3.042	14.344 ± 3.027	14.350 ± 3.020
TV ($\alpha=9e-4$)	14.311 ± 2.821	14.379 ± 2.820	14.420 ± 2.808	14.496 ± 2.828	14.510 ± 2.819	14.523 ± 2.819
TV ($\alpha=5e-4$)	14.223 ± 2.799	14.306 ± 2.803	14.376 ± 2.798	14.448 ± 2.815	14.468 ± 2.809	14.493 ± 2.811
TV ($\alpha=1e-4$)	14.114 ± 2.769	14.215 ± 2.781	14.325 ± 2.786	14.392 ± 2.799	14.420 ± 2.796	14.460 ± 2.802

tree anatomy with our non-optimized implementation are shown in Table 3.9 as measured on a local workstation using 8 CPUs. For the SRR obtained by NiftyMIC using the 'a+c+s+3obl' as source data configuration results in a total computational time of about 40 min using a TK1 regularizer. Using TV instead increases the total computational time by nearly 300 % to about 2 h 30 min.

Fast SRR processing times are not a main consideration in the context of obtaining high-resolution visualizations of abdominal anatomy from MRCP data. However, our results underline that TV regularization substantially increases the computational cost but tends to show only little improvement in the obtained reconstruction quality.

3.4.3. Discussion

In this section, we presented the first comprehensive analysis of SRR for MRCP studies. Using quasi-static control data from the brain and upper abdominal MR imaging

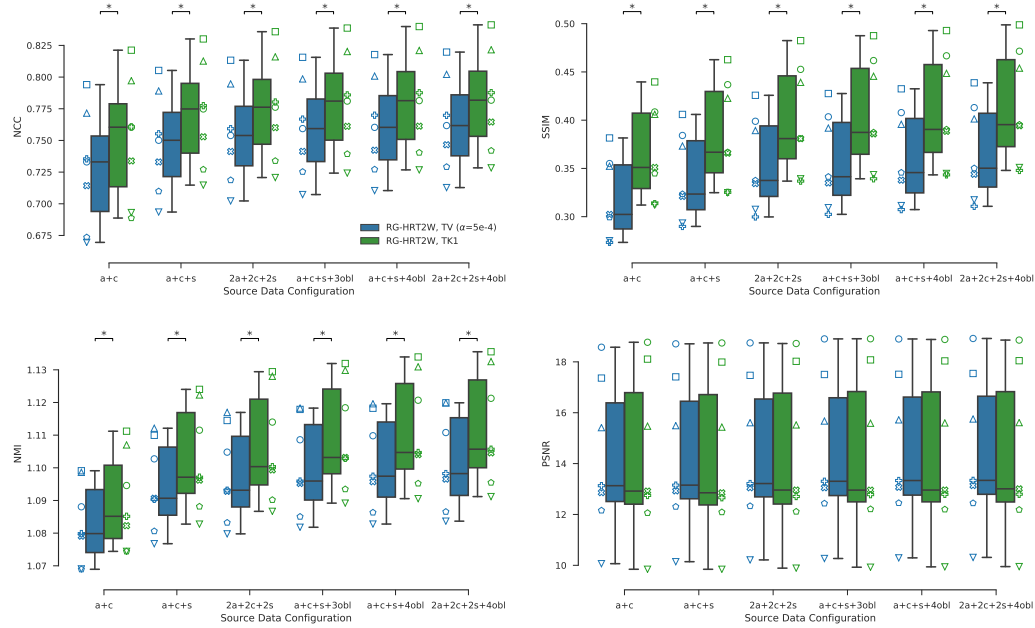


Figure 3.13.: Ground-truth (HR T2W) similarities for first-order Tikhonov (TK1) and isotropic Total Variation (TV) regularization SRR outcomes for the quasi-static brain experiment involving seven subjects. The respective regularization was applied in the final reconstruction step using RG-HRT2W-NCC as the motion-correction strategy. Stars indicate statistical differences between the groups using a pairwise Wilcoxon signed-rank test ($p < 0.05$).

of healthy volunteers, we optimised source data configuration and motion-correction strategies for MRCP SRR. We validated optimised SRRs using expert-readers to show that SRRs can lead to novel high-quality HR visualization of peri-ductal anatomy. In particular, we empirically showed that SRRs based on the clinically available axial and coronal images are of inferior quality compared to those with additional input stacks. By performing highly controlled brain experiments we found that after approximately six input stacks the SRR quality achievable with MRCP SST2W sequences plateaus. Moreover, we showed that not only the number but also the orientation of the SST2W stack acquisitions matters. In particular, for the same number of six input stacks using oblique orientations on top of the standard anatomical directions produces superior SRR outcomes compared to using multiple standard axial, coronal and sagittal anatomical acquisitions. Notably, motion correction was needed for these experiments to exclude the confounding factor of motion despite the 'static'

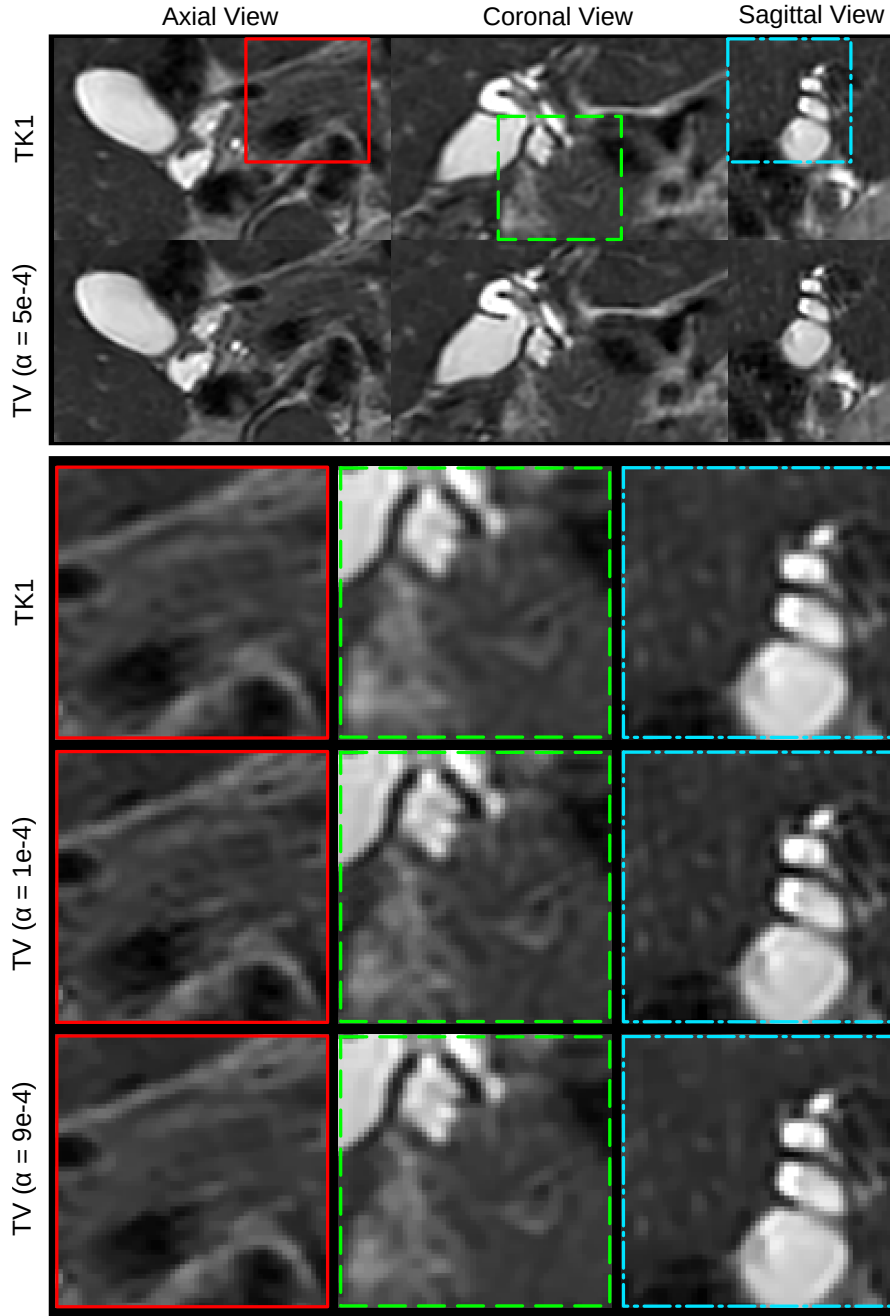


Figure 3.14.: Qualitative comparison between first-order Tikhonov (TK1) and isotropic Total Variation (TV) regularization in the final reconstruction step using NiftyMIC (a+c+s+3obl). The zoomed windows illustrate that smaller regularization parameters α for TV result in similar SRRs as obtained by TK1. Increasing α leads to reconstructions with slightly sharper edges but at the cost of a staircasing effect typical for TV regularization [Chambolle and Pock, 2011] which presents artificial discontinuities and may suppress clinically relevant structural information.

Table 3.9.: Typical computational times to create a HR visualization of the biliary tree split into motion correction and volumetric reconstruction processing times. Motion correction for NiftyMIC refers to the total time of the two-step registration and TK1-based reconstruction iterations without the final SRR step. Volumetric reconstruction refers to solving the SRR problem (3.2) to obtain an isotropic volume of 0.78 mm resolution after performed motion correction using either first-order Tikhonov (TK1) or isotropic total variation (TV) regularizations (3.5) and (3.6), respectively. The total reconstruction time is determined by the sum of the individual motion-correction and volumetric reconstruction times.

Source Data Configuration	Motion Correction:		Volumetric Reconstruction:	
	RG-HT2W-NCC	NiftyMIC	TK1	TV
a+c	1 min 30 s	7 min 20 s	1 min	17 min
a+c+s	2 min 30 s	10 min 40 s	1 min 30 s	25 min
a+c+s+3obl	7 min 15 s	32 min 15 s	7 min 15 s	2 h

nature of brain imaging.

High anatomical fidelity of the SRRs relies on the accurate establishment of generally non-linearly affected, anatomical correspondences captured by different SST2W stacks acquired at different breath-holds. We explored two SRR frameworks that are based on different motion-correction strategies: A non-iterative framework based on reference-guided multimodal, in-plane deformable motion correction that leverages the existence of a separate HR volume of a different modality for motion correction, and an iterative outlier-robust SRR framework that is based on monomodal rigid motion correction that does not rely on any external reference. For the reference-guided approach, we found that the optimal combination of reference image and associated similarity measure for registration is difficult to determine reliably. Whereas the controlled quasi-static brain experiments indicated high quality SRR outcomes by using the BFFE combined with NCC, this set-up failed in the abdominal experiments due to misregistrations. We hypothesize that this comes from more complex appearance differences between BFFE and SST2W image contrasts we noticed in the abdomen compared to the brain, including the more pronounced cancellation effects around fat-water tissue boundaries typical for BFFE (Figures 3.6 and 3.15). Besides the reported similarity measures and references in here, we also ran experiments using NMI, MI and LNCC as additional similarity measures and the standard axial SST2W series as alternative reference, all of which corroborated the findings

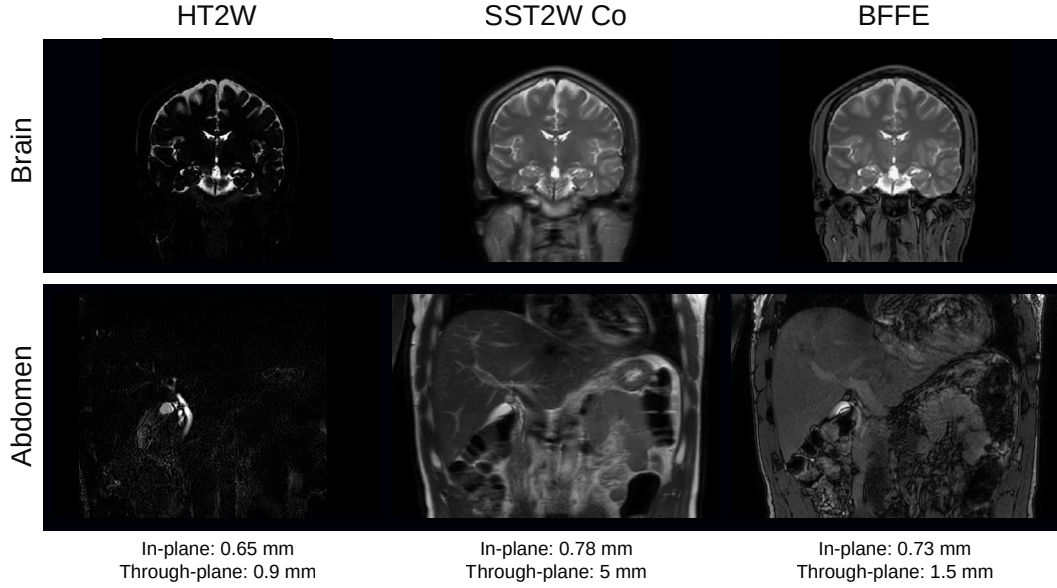


Figure 3.15.: T2W brain and abdominal data associated with the acquisition protocol defined in Table 3.2. For the abdomen, more complex appearance differences between BFFE and SST2W image contrast can be noticed, including the more pronounced cancellation effects around fat-water tissue boundaries typical for BFFE.

using NCC (Tables B.1 and B.3). Among reference-guided SRR frameworks applied to the abdomen, only the HT2W-guided approach driven by NCC showed promising results in regions of high MRCP contrast but is prone to artefacts in areas where this is not the case. Our results indicate that the RG-HT2W-based motion correction can lead to improved anatomical clarity compared to the Static SRR for the abdomen. However, it is prone to creating artefacts due to slice misregistrations which may degrade the overall reconstruction quality. Additional experiments suggest that the in-plane deformation step does not substantially improve the results (Table B.3). Similarly, experiments incorporating the PSF for registration remained inconclusive (Figure 3.12 and Tables 3.5 and 3.6). Contrastingly, we found very encouraging outcomes for the outlier-robust SRR approach that was originally developed for fetal brain MRI. Despite the use of a rigid motion-correction model only, it could consistently generate SRRs of the biliary tree that have the potential for diagnostic use. We therefore conclude that the encountered motion was approximately rigid for the most part and that the outlier rejection mechanism is effective in eliminating slices

where this was violated. In case of sufficient data oversampling, the SRR algorithm was then able to reconstruct anatomical structures with high anatomical clarity. Qualitative expert-reader comparisons showed that the optimised SRRs tend to provide limited value for regions like ampulla, the head of pancreas parenchyma, the imaged pancreatic duct and the CBD where specifically developed MRCP SST2W sequences traditionally show high diagnostic yield. However, in regions where the SST2W data typically provides only limited insight such as the portal vein and first generation branches, hepatic artery and central branches, cystic duct, and the imaged first generation intrahepatic ducts the SRRs allow an assessment with much higher anatomical detail which has important applications particularly for the assessment of hepato-pancreatico-biliary cancers. Therefore, we believe that using SRRs alongside the original data has real potential to improve the diagnostic yield of standard MRCP imaging, and improve patient care by reducing delays introduced by the need for further investigations, particularly in the context of cancer care.

Limitations of this work include the analysis of a relatively small cohort of eight subjects. Moreover, this pilot study was conducted using healthy volunteers. Acquired data during breath-holds of patients is expected to have more challenging motion artefacts. We therefore plan to apply the proposed extended MRCP protocol to a representative patient cohort to assess the clinical utility of MRCP SRR. Furthermore, we plan to make improvements on the reference-guided framework. Using the BFFE as a reference appears promising in case a more robust similarity measure is available. In fact, recently proposed deep learning methods [Cheng et al., 2018, Simonovsky et al., 2016] could prove useful for this step or, as shown in other applications, be used to aid the motion correction as a whole [Hu et al., 2018]. In this work, we performed manual segmentations to define the region of interest in the standard axial SST2W image for the SRR. This step could be automated similar to the work as proposed in, e.g., [Ebner et al., 2018c]. For NiftyMIC a unified motion-correction/reconstruction framework could help to better constrain the registration steps that might also allow the incorporation of a deformable model. This could help achieve higher anatomical accuracy in correcting for the challenging deforma-

tions in the upper GI anatomy. Finally, validation with abdominal isotropic 3D volumes acquired as “single-slab” 3D fast-spin-echo sequences [Mugler, 2014], was not undertaken in this study, but remains subject of future work.

CHAPTER 4.

Volumetric Reconstruction from Printed Films: Enabling 30 Year Longitudinal Analysis in MR Neuroimaging

Table of Contents

4.1. Introduction	155
4.2. Volumetric Reconstruction from Printed MR Films . .	161
4.2.1. Semi-Automatic Slice Extraction	164
4.2.2. Joint Regularised Motion and Intensity Correction Model .	165
4.2.3. Initialisation of Volumetric Reconstruction Algorithm . . .	166
4.2.4. Global Intensity Correction	167
4.2.5. Refinement of Volumetric Reconstruction	167
4.3. Data, Evaluation Methodology and Results	169
4.3.1. Data	169
4.3.2. Parametrisation of Volumetric Reconstruction Pipeline . . .	170
4.3.3. Evaluation Methodology	172
4.3.4. Results	174
4.4. Discussion	181

Foreword. This chapter is adapted and contains content from the work published in

Ebner, M., Chung, K. K., Prados, F., Cardoso, M. J., Chard, D. T., Vercauteren, T., and Ourselin, S. (2018a). Volumetric reconstruction from printed films: Enabling 30 year longitudinal analysis in MR neuroimaging. *NeuroImage*, 165:238–250.

My specific contributions presented in this chapter are:

- A publicly available¹ volumetric reconstruction framework to recover geometrically consistent, volumetric images from printed films in MR neuroimaging
- The presented framework consists of
 - a semi-automatic slice extraction tool to create a digital image stack from historical slices selected from scanned brain MR films, and
 - a fully automatic volumetric reconstruction framework to estimate the lost meta-data information of each slice in the 3D space.
- Methodological contributions include
 - a reconstruction framework based on a joint slice-to-volume affine registration with inter-slice 2D transformation regularisation and affine slice-intensity correction whereby the missing meta-data information is contributed by a longitudinal scan of the same subject
 - a final isotropic total variation in-plane deconvolution technique serves to revitalise the visual appearance of the reconstructed stack of historical slices

For a follow-up work [Chung et al., 2019], I used the presented framework to reconstruct more than 200 image acquisitions captured on historical films to support a landmark longitudinal study² of 132 multiple sclerosis (MS) patients which were tracked since the 1980s. These reconstructions may unlock the full potential of modern image analysis tools and help understand the disease progression in MS.

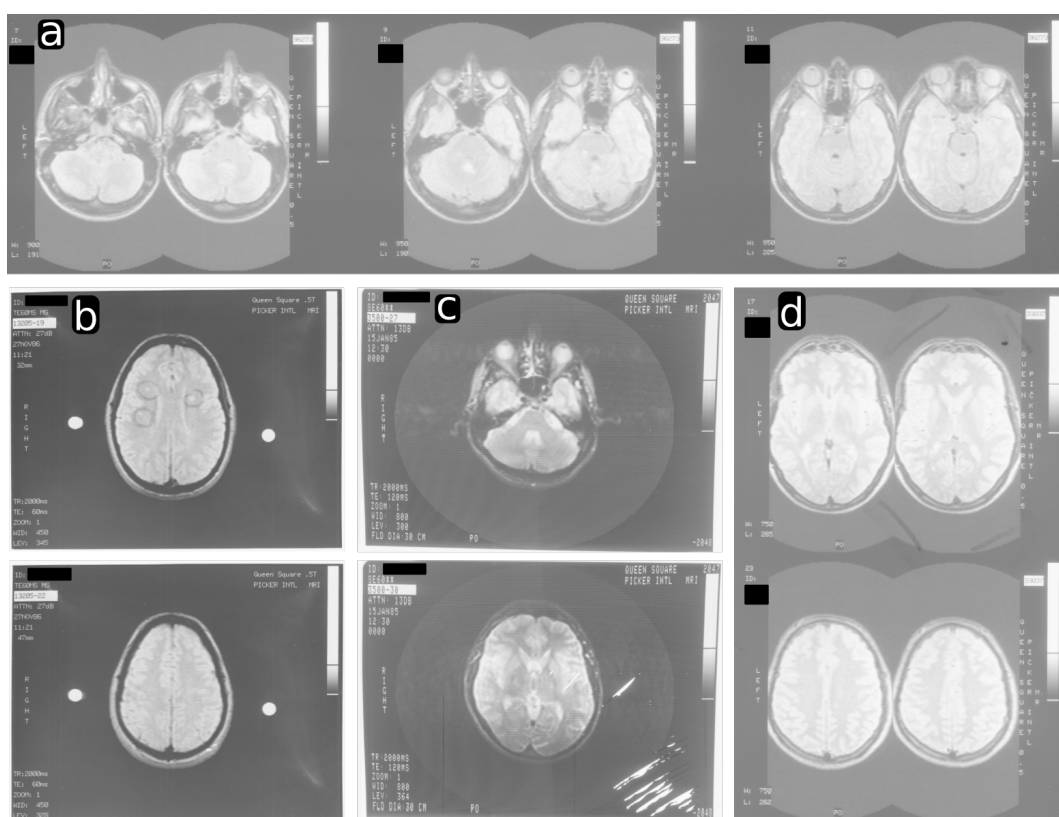


Figure 4.1.: Scanned MR Films of MS/Clinically Isolated Syndrome (CIS) subjects from 1985 to 1991. The top row (a) illustrates a typical proton density-like sequence of printed 5 mm thick slice acquisitions side-by-side, acquired at 1.5 T in 1991. The left bottom column (b) illustrates a scan acquired at 0.5 T in 1986 with lesions encircled directly on the film. The middle bottom column (c) depicts a scratched, visibly rotated and deformed film from 1985 with unknown slice-thickness. The bottom right column (d) shows handwritten annotations on a scan from 1990. The skulls and also a part of the brain is merged into each other to save film space. Overall, it is worth noting the different types of MR films, their illumination differences and different visible distortions affecting even the same film belonging to the same acquisition.

4.1. Introduction

As pointed out in Chapter 1, Magnetic Resonance Imaging (MRI) has been recognised as a powerful, non-invasive and non-ionising medical imaging technique since the early 1980s, when it first became available for clinical use [Damadian, 1971].

¹<https://github.com/gift-surg/VolumetricReconstructionFromPrintedFilms>

²<https://www.ucl.ac.uk/healthcare-engineering/news/2018/feb/image-reconstruction-tool-allows-researchers-unlock-historic-ms-data>

The earliest, longitudinal brain studies were performed based on thick contiguous slices acquired in the axial direction to cover the entire volume, e.g. [Miller et al., 1988, Miller et al., 1989]. In absence of modern standards for digital archives and visualisation, the acquired scans were placed side-by-side and printed sequentially on multiple films, for further, *visual* analysis, as shown in Figures 1.3 and 4.1. The analysis was typically limited to measures such as lesion count and location in multiple sclerosis (MS) studies [Miller et al., 1989, Morrissey et al., 1993, O’Riordan et al., 1998]. With the introduction of the PACS (Picture Archiving and Communications System) and DICOM (Digital Imaging and COmmunication in Medicine) standards in the beginning of the 1990s, the basis was created to digitally store medical imaging information including essential meta-data on spatial information and acquisition details. This allowed further development of clinically important biomarkers such as brain and lesion volume for longitudinal MS studies [Brex et al., 2002, De Stefano et al., 2014, Sailer et al., 1999] – information which, currently, cannot be readily extracted from scans dating back to the 1980s and early 1990s if they are only available as printed films. The original, digital data is often lost or cannot be recovered due to hardware and software obsolescence issues which has also been shown in more recent studies on brain morphometry where original MR films were digitised anew and manually processed to allow for further quantification [Ekert et al., 2016]. In other words, especially for longitudinal studies dating back to the 1980s, *a decade or more of valuable image data information may not be readily accessible to modern image processing techniques* which could add to the understanding of long-term pathological or morphological evolution.

In order to re-establish a consistent, volumetric representation from printed, historical films several challenges need to be overcome: Patient-specific anatomy is only sparsely captured on printed films corresponding to a single acquisition of axially acquired thick slices only. Similar to the applications developed in Chapters 2 and 3, each 2D slice needs to be extracted from the MR film and correctly aligned in the 3D space. Slice-based motion correction and reconstruction frameworks were proposed for structural MRI (see fetal and abdominal applications as described earlier in Chap-

ters 2 and 3), but also for functional MRI [Kim et al., 1999, Seshamani et al., 2016], diffusion tensor imaging (DTI) [Jiang et al., 2009, Fogtmann et al., 2014, Marami et al., 2016, Marami et al., 2017] and histology [Casero et al., 2017]. In particular, [Fogtmann et al., 2014] and, more recently, [Marami et al., 2017] have proposed regularised slice-to-volume registration approaches to better address the ill-posed nature of independent slice-to-volume registrations in order to achieve more robust motion correction frameworks for *multi-plane multi-slice* Diffusion MRI. In the case of reconstructing volumetric representations from printed MR films, however, only one *single* stack of past thick-slice acquisitions in a *single, axial plane* is captured. The slice thickness of past acquisitions can range from 5 mm like in [Brex et al., 2002, Ekert et al., 2016, Miller et al., 1989, Sailer et al., 1999] to 10 mm as in [Brex et al., 2002, Miller et al., 1989, Sailer et al., 1999] to even encountered 12 mm. Hence, even in the 5 mm slice thickness case, which is the focus of this work, neighbouring slices correspond to relatively distant anatomical positions which renders purely intra-stack alignment-based motion correction approaches particularly difficult so as to recover the correct inter-slice relationship and, thus, the patient-specific anatomy. This is in contrast to the method presented in [Casero et al., 2017] for histology where purely intra-stack alignment steps were performed for the 10 μ m thick slices after completing a first initial alignment using a lower-resolution blockface image as reference [Casero et al., 2016, Siedlecka et al., 2013]. An appropriate single slice-based motion correction approach will be key but needs to deal with the very sparse information given the thick, contiguous slices. Additionally, the geometrical properties and dimensions of printed slices are lost and need to be recovered. The top part of the brain is often missing due to a reduced field of view (FOV) in past acquisitions complicating accurate registrations. The arguably higher magnetic field inhomogeneities of past MR image acquisitions and the further processing associated with film printing, storage and subsequent scanning lead to different types of illumination differences which are present across, but also within, MR films belonging to the same acquisition in addition to other types of degradations as shown in Figure 4.1. Storage of the films over decades may have further degraded the data whereby individual films belonging to

the same acquisition may have been affected differently resulting in stark differences in image intensities across slices of different films. Moreover, historical films are likely to carry a substantial amount of background noise and may well come with low image contrast. Additional distortion has been introduced due to the performed manual scanning, manifested in rotated, sheared and possibly, otherwise deformed images, as visible in Figure 4.1. Due to advances in MR, increased field strengths, higher spatial resolution, changes in imaging protocols and image contrast preferences for diagnostic purposes in addition to changes in MR scanner manufacturers and printers, the appearance and also the layout of MR films can change substantially in the course of a longitudinal study spanning several decades. In particular, this renders learning-based registration approaches such as presented in [Miao et al., 2016, Hou et al., 2018] difficult as not many data samples are available for the training stage.

In this chapter, we propose a novel reconstruction framework, able to address the challenges discussed above. More specifically, our contributions are:

1. A semi-automatic slice extraction tool to create a digital image stack from historical slices selected from the scanned brain MR films. It provides an initial digital 3D representation of acquired slices printed on a 2D film where the correct spatial position and dimension of each single slice needs to be recovered.
2. A fully automatic volumetric reconstruction framework to estimate the lost meta-data information of each slice in the 3D space. It is based on a joint slice-to-volume affine registration with inter-slice 2D transformation regularisation and affine slice-intensity correction. Missing meta-data information is contributed by a longitudinal scan of the same subject. A final isotropic total variation in-plane deconvolution technique serves to revitalise the visual appearance of the reconstructed stack of historical slices.
3. A validation of our slice-extraction tool and volumetric reconstruction framework on clinical, *historical ground-truth* data to show the potential of our proposed framework to enable a more robust analysis of long-term datasets:

- We apply our proposed method to a uniquely long-term, longitudinal dataset of patients first recruited with clinically isolated syndrome (CIS) dating back to the 1980s [Miller et al., 1988, Miller et al., 1989, Morrissey et al., 1993, O’Riordan et al., 1998, Brex et al., 2002, Fisniku et al., 2008].
- We validate our framework on a subset of this cohort where also the original, digital stack of the same acquisition has been preserved in addition to the printed MR films. In this rare situation, we can validate against *historical ground-truth* data.
 - We perform a quantitative comparison and assess the accuracy of our obtained volumetric reconstructions in terms of mean squared error, normalised cross correlation, structural similarity, peak signal-to-noise ratio and Structural Image Evaluation, using Normalization, of Atrophy (SIENA) [Smith et al., 2002] analysis.
 - We undertake a qualitative assessment relying on expert neurologist ratings both in terms of clinical usefulness and ground-truth comparison of our recovered volumetric representations of historical film data.
 - We provide a qualitative comparison of longitudinal data to assess ground-truth similarity over time.

The framework is made open source and available on github³.

Compared to regularised slice-to-volume motion-correction and MR reconstruction methods proposed in the literature such as [Fogtmann et al., 2012, Fogtmann et al., 2014] or [Marami et al., 2017] the proposed reconstruction pipeline differs significantly in a number of aspects as it is designed particularly to deal with the specific challenges associated with the volumetric reconstruction from historical MR films.

[Fogtmann et al., 2014] propose reconstructing 3D DTI from *multiple* multi-slice diffusion weighted (DW) images by using a framework for unified motion-estimation and image reconstruction as an extension of their previous work [Fogtmann et al.,

³<https://github.com/gift-surg/VolumetricReconstructionFromPrintedFilms>

2012] introduced for structural *multi-plane* MRI. Despite the formulation as a unified approach, the volumetric reconstruction of the unknown image is performed by alternating between the two problems of estimating the motion parameters of a 3D rigid and scale-skewness transform for all slices followed by estimating the weight parameters which define the diffusion volume. Instead, we propose a method which corrects at once for affine in-plane 2D motion of each single slice and estimates its missing physical dimensions by the guidance of a reference volume acquired many years later which usually exists in longitudinal studies. In particular, this approach avoids out-of-plane resampling of the very sparse anatomical data given by only one *single* stack of the thick axial slices. Moreover, [Fogtmann et al., 2014] use a regularisation prior based on the Huber norm for motion correction to favour similarity between motion correction parameters of consecutive slices. We propose a robust smooth ℓ^1 -approximation-based inter-slice affine 2D transform regularisation and affine intensity correction framework in addition to the use of a prior on optimisation parameters based on a smooth ℓ^1 -approximation. This drives the physical dimension and 2D position estimates directly by the similarity between slice neighbours guided by the anatomical shape prior provided by the reference volume.

[Marami et al., 2017] build on their work presented in [Marami et al., 2016] and explicitly model the dynamics of rigid motion with a state space model where they estimate the temporal motion trajectories with a Kalman filter for a more robust reconstruction of DWI. They automatically detect and reject motion-corrupted DWI slices to enhance motion tracking and reconstruction. In our setting of volumetrically reconstructing 3D volumes from sparse historical 2D slices printed on films, the motion captured in the obtained stack of slices after the semi-automatic slice extraction cannot be assumed to follow a physiological model. Each individual slice will have different motion shifts with respect to each other given that each slice is extracted according to a landmark which is placed manually on the film. Therefore, we propose a motion correction framework based on an inter-slice regularisation which leverages the 2D image similarity between two neighbour slices and the respective (oblique) reference slice instead. Moreover, we do not perform outlier rejection as we

need to recover the physical position and dimension for each single slice reliably in order to form a consistent volumetric representation of the subject-specific anatomy as captured by the single acquisition in the past. Importantly, however, [Marami et al., 2017] use the method presented in [Kainz et al., 2015b] for the required structural image reconstruction of the high-resolution T2-weighted volume from multiple motion corrupted slices, which uses independent rigid slice-to-volume registration without regularisation.

In addition, the mentioned approaches rely on Super-Resolution techniques (see Chapters 2 and 3) for the image reconstruction steps to reconstruct a single, higher-resolution, isotropic 3D volume from multiple scattered low-resolution 2D slices. In the approach developed in this chapter, a final isotropic total variation in-plane deconvolution step is added after performed in-plane motion correction of each single slice for improved visual appearance only.

The remaining part of this chapter is organised as follows. Section 4.2 motivates and presents the design choice and the details of our proposed volumetric reconstruction framework. In Section 4.3 the results of the validation of our proposed volumetric reconstruction framework are presented using a uniquely long-term historical dataset spanning 30 years of MRI scans. Finally, Section 4.4 concludes with a discussion.

4.2. Volumetric Reconstruction from Printed MR Films

The first step of the volumetric reconstruction method is dedicated to the slice extraction and stacking of all slices of the same historical axial acquisition in order to create an initial, digital 3D image. A semi-automatic slice extraction framework is chosen to deal with the wide variety of existing films, data characteristics, and styles as pointed out in Figure 4.1. This enables flexible processing even for complex cases where slices are merged on a printout and brain images need to be carefully delineated, as, e.g., shown in Figure 4.1a. A manual interaction can additionally ensure that only correct slices and films belonging to the same acquisition are extracted. This is particularly relevant since multiple films, or slices printed on films, encoun-

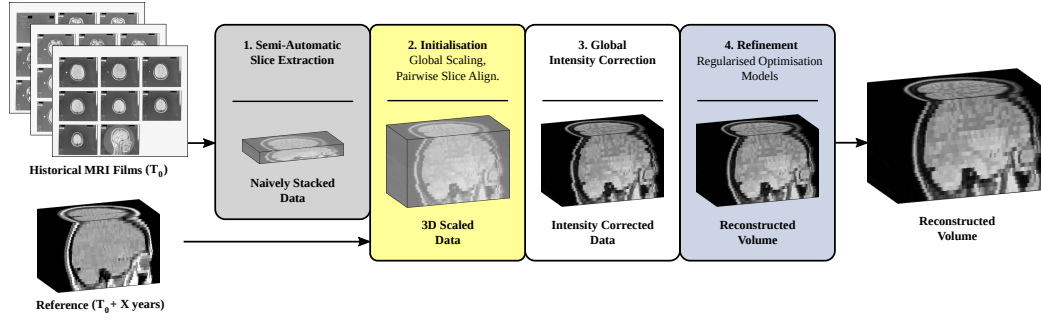


Figure 4.2.: Overview of volumetric reconstruction framework for historical MR film data. Provided the scans of the MR films acquired at time T_0 and a more recent, digital 3D scan of the same patient acquired X years later, the proposed algorithm reconstructs the volumetric representation of the original MR image acquisition at T_0 . Further details on the semi-automatic slice extraction, initialisation and refinement steps are visualised in the Figures 4.3 to 4.5, respectively.

tered in the database of historical films are duplicated whereas other ones are missing or not ordered in the right sequence and inevitably require manual intervention.

The imperfect slice extraction of rotated, sheared and possibly otherwise deformed images due to printing and manual scanning gives rise to a naively stacked 3D data with visibly in-plane motion affected slices. Therefore, the volumetric reconstruction framework needs to recover the correct inter-spatial position of all slices in addition to their physical dimensions. After printing and scanning, the only spatial information available from the MR films is the slice thickness which is generally printed on the films as indicated in Figure 4.1. All other meta-data typically stored in DICOM headers, such as exact spatial position relative to the neighbouring slice, in-plane spacing, and image orientation, is lost and needs to be recovered for each single slice. The aim of our algorithm is to infer the missing information from a more recent, digital 3D scan of the same patient which holds the required meta-data information and is of similar intensity contrast; a scan which generally exists in longitudinal studies. For instance, early and current studies for MS use proton density (PD)-like image contrast [Gracien et al., 2016, Miller et al., 1988] despite the advances and changes in imaging protocols over several decades. However, this reference image is likely to be acquired many years later and the patient may have undergone substantial morphological changes including atrophy. It will be dissimilar to the brain captured

by the historical MR films.

By neglecting subject motion during acquisition time, slice motion correction can be reduced to in-plane motion correction only. This assumption addresses both the need of balancing the complexity versus robustness of the method and avoids out-of-plane resampling for the final volumetric reconstruction of the sparse, historical, thick-sliced data. The primary goal of our obtained volumetric reconstructions from printed films is to gain clinical trust by performing a sufficiently accurate motion correction without introducing implausible deformations. Hence, a gradual increase of transformation complexity shall be performed up to in-plane 2D affine transformation which is believed to be capable of dealing with the encountered distortions in the films. Intra-stack slice registration is highly ill-posed due to the large slice thickness and its associated sparse anatomical sampling of the patient-specific brain anatomy. This inhibits a potential approach of performing first an intra-stack slice alignment followed by a subsequent volume-to-volume registration to the reference image. Taking advantage of the valuable information on the skull geometry captured by the later reference image with similar appearance, we propose leveraging the combined information of both reference and historical slice neighbour data by deploying a slice-to-volume registration framework based on regularised motion and affine intensity correction. The regularised slice-based intensity correction is meant to deal with the intensity variations across slices and films and to balance existing intensity discrepancies to the reference image. A previous *global* intensity correction step will be vital to eliminate background noise of the historical slices and to scale the scanned image intensities accordingly. Finally, we make use of a well-established, isotropic total variation deconvolution step [Beck and Teboulle, 2009b, Rudin et al., 1992] to alleviate the blurring of the historical slices resulting from both printing and scanning steps and to reduce the impact of the original point-spread-function (PSF) from the MR scanner during acquisition time.

Based on those assumptions, the proposed algorithm, illustrated in Figure 4.2, reconstructs the 3D geometry of the original shape as captured by the MR films such as shown in Figure 4.1.

4.2.1. Semi-Automatic Slice Extraction

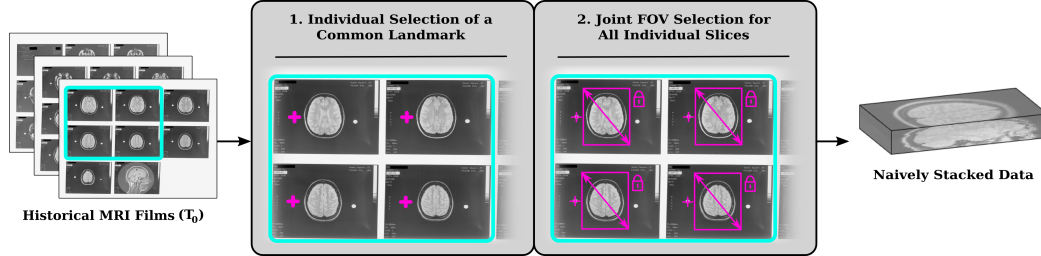


Figure 4.3.: Semi-Automatic Slice Extraction. By providing the scanned MR films associated with the same acquisition a semi-automatic slice extraction step is deployed to naively stack each extracted 2D slice into a 3D stack. No meta-data information is assigned at this stage.

A flexible semi-automatic procedure is proposed to extract each individual slice-acquisition from every MR film associated with the same acquisition to address the variety of existing films and styles as pointed out in Figure 4.1. As shown in Figure 4.3, the first MR film is read and a common landmark is selected manually with one click per slice on the film. The selection order of the landmarks defines the slice extraction order for the later slice stacking. This ensures that the proposed slice extraction tool can deal with historical MR films where consecutive acquisitions are not necessarily printed sequentially and in the same manner across different film types. We emphasise that the manual landmark selection does not need to be very precise and is used for initialisation purposes of the motion correction algorithm only. After landmark selection, a selection window is automatically overlaid, based on default values relative to each landmark, indicating the FOV for slice extraction. The window size and respective offset are then adjusted manually, in a uniform manner, so that changes applied to one window are automatically adjusted to the rest. A more precise extraction window is then achieved, comprising the region of interest for all slices on the film. This adjustment also allows to easily extract slices even in cases where skulls are overlapping as shown in Figure 4.1a. The final FOV windowing is stored and used for the subsequent films. After having marked the same common landmark on all slices on all remaining films sequentially, the selected 2D slices are extracted and stacked automatically to form a naively stacked 3D data.

4.2.2. Joint Regularised Motion and Intensity Correction Model

We propose using an in-plane, affine spatial transformation model to strike a balance between fully compensating the distortions of each slice while preventing the introduction of additional image artefacts. However, due to the sparse, thick slice anatomical sampling, additional information is required to estimate the true, original, anatomical shape. The missing information can be contributed by the reference image. Due to the mentioned morphological changes of this later scan, we will use only the information around the skull; the structure which is believed to change the least over time. Hence, a good estimate of the correct slice position and geometry is likely to be found if each historical slice y_k matches both a corresponding reference (oblique) slice r_k around the skull and its immediate neighbouring slices over the FOV to achieve a good geometric consistency driven by the original MR film data. Thus, the idea is to find the slice transformation parameters $\boldsymbol{\theta}_k \in \mathbb{R}^6$ to an in-plane 2D affine spatial transformation $\mathbf{T}(\boldsymbol{\theta}_k, \cdot) : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}^2$ which jointly minimises the costs

$$s_1\left(y_k(\mathbf{T}(\boldsymbol{\theta}_k, \boldsymbol{\xi})), r_k(\boldsymbol{\xi})\right) \quad \forall \boldsymbol{\xi} \in \Omega_{\text{skull}} \subset \Omega, \quad (4.1)$$

$$s_2\left(y_k(\mathbf{T}(\boldsymbol{\theta}_k, \boldsymbol{\xi})), y_{k+1}(\mathbf{T}(\boldsymbol{\theta}_{k+1}, \boldsymbol{\xi}))\right) \quad \forall \boldsymbol{\xi} \in \Omega, \quad (4.2)$$

for adequate similarity measures s_1, s_2 for all K slices. To compensate intensity variations across historical slices and between historical and reference slice, a global intensity correction step, as shown in Figure 4.2, and a local affine intensity compensation model during registration will be deployed. By defining the joint motion and intensity correction parameter $\boldsymbol{\Theta}_k := (\boldsymbol{\theta}_k, \alpha_k, \beta_k)$ the cost (4.2), measuring the dissimilarity of neighbouring, historical slices, is defined as

$$\begin{aligned} \mathcal{N}_k(\boldsymbol{\Theta}_k, \boldsymbol{\Theta}_{k+1}) := & \sum_{\boldsymbol{\xi} \in \Omega} \varrho_\gamma \left(\alpha_k y_k(\mathbf{T}(\boldsymbol{\theta}_k, \boldsymbol{\xi})) + \beta_k \right. \\ & \left. - \alpha_{k+1} y_{k+1}(\mathbf{T}(\boldsymbol{\theta}_{k+1}, \boldsymbol{\xi})) - \beta_{k+1} \right) \end{aligned} \quad (4.3)$$

with a loss function $\varrho_\gamma : \mathbb{R} \rightarrow \mathbb{R}^+$, $e \mapsto \varrho_\gamma(e) := \sqrt{\gamma^2 + e^2} - \gamma$, as a smooth ℓ^1 -approximation with scaling factor $\gamma > 0$, similar to the Huber function, to allow for a more robust optimisation. Similarly, the cost (4.1) between historical and corresponding reference slices is defined as

$$\mathcal{R}_k(\boldsymbol{\Theta}_k) := \sum_{\boldsymbol{\xi} \in \Omega_{\text{skull}}} \varrho_\gamma(\alpha_k y_k(\mathbf{T}(\boldsymbol{\theta}_k, \boldsymbol{\xi})) + \beta_k - r_k(\boldsymbol{\xi})) \quad (4.4)$$

whereby only a neighbourhood around the skull $\Omega_{\text{skull}} \subset \Omega$ is considered. With $\boldsymbol{\Theta} := (\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_K)$ denoting the joint set of optimisation parameters and $\mathcal{P}(\boldsymbol{\Theta}) := \sum_{\theta \in \boldsymbol{\Theta}} \varrho_\gamma(\theta)$, the motion correction problem for one stack of semi-automatically extracted slices can subsequently be written as a joint, regularised minimisation problem

$$\min_{\boldsymbol{\Theta}} \left(\lambda_{\mathcal{R}} \sum_{k=1}^K \mathcal{R}_k(\boldsymbol{\Theta}_k) + \lambda_{\mathcal{N}} \sum_{k=1}^{K-1} \mathcal{N}_k(\boldsymbol{\Theta}_k, \boldsymbol{\Theta}_{k+1}) + \lambda_{\mathcal{P}} \mathcal{P}(\boldsymbol{\Theta} - \boldsymbol{\Theta}_0) \right) \quad (4.5)$$

with weights $\lambda_{\mathcal{R}}, \lambda_{\mathcal{N}} > 0$, regularisation parameter $\lambda_{\mathcal{P}} > 0$ and prior $\boldsymbol{\Theta}_0$ on the parameters which need to be defined accordingly. The critical point is to get the corresponding reference (oblique) slices r_k , $k = 1, \dots, K$, to initialise (4.5).

4.2.3. Initialisation of Volumetric Reconstruction Algorithm

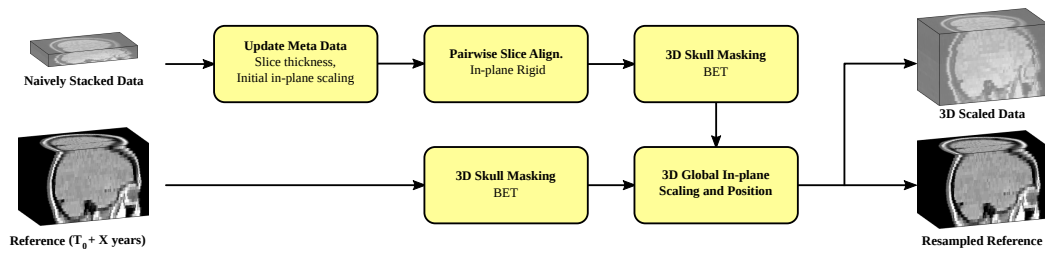


Figure 4.4.: Initialisation of the volumetric reconstruction algorithm: Global rescaling and positioning of the historical slices to initialise the regularised optimisation models.

As mentioned earlier, the slice thickness is the only preserved meta-data information of the slice acquisitions stated on the MR films. After the scanning, the image data is only given in pixel dimensions. Hence, the semi-automatically extracted and

naively stacked slices need to be re-scaled and aligned with the reference 3D volume in order to extract the respective reference (oblique) slices r_k . Our proposed approach is shown in Figure 4.4. The slice-thickness is updated according to the information from the MR film and the in-plane scaling is initialised by a uniform value manually fixed. A subsequent rigid in-plane registration step with inter-slice regularisation is performed using the sum of the slice neighbour-terms (4.3) to obtain a more consistent brain geometry and to correct for possible inaccuracies of the semi-automatic slice extraction. More information on the parametrisation of the registration parameters is provided in Section 4.3.2. By using the Brain Extraction Tool (BET) [Smith, 2002] a mask surrounding the skull of both the in-plane registered stack and the reference image is created. A subsequent in-plane 3D similarity registration based on normalised cross-correlation aligns the entire stack with the reference and estimates a uniform in-plane scaling factor for all slices to match their skull masks. The additionally resampled reference image to the 3D scaled stack space both provides the oblique slices r_k and serves for the subsequent global intensity correction step.

4.2.4. Global Intensity Correction

Due to the printing, storage over years and scanning, the historical slices may carry a substantial amount of background noise and have low image contrast. A global intensity correction step aims to improve the intensity contrast by using the re-sampled 3D reference image intensity information and to keep the slice intensities as close as possible to the original ones' at the same time. With $q_{20\%}$ being the global 20 %-intensity percentile of all historical slices, all slice intensities i are capped via $i \leftarrow \max(i - q_{20\%}, 0)$ to eliminate background noise whereby the 20 %-threshold was found experimentally by visual analysis. We then apply a uniform-across-slices linear intensity correction step.

4.2.5. Refinement of Volumetric Reconstruction

In order to increase the convergence basin of the joint-regularised registration model (4.5), a step-wise increase in transformation complexity is chosen for the **slice-based mo-**

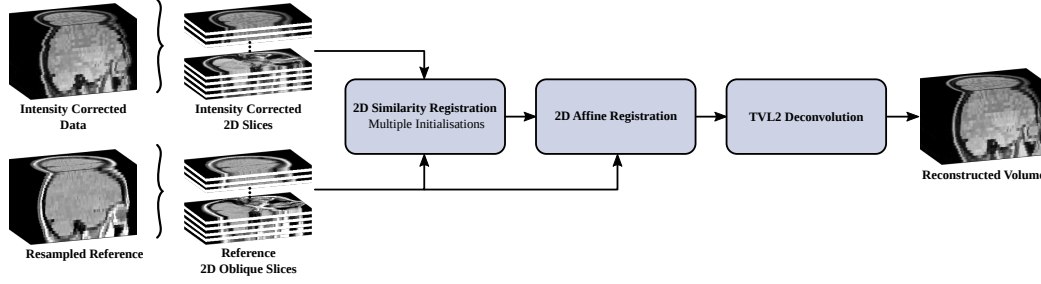


Figure 4.5.: Refinement of the volumetric reconstruction algorithm: Application of joint, regularised motion and intensity correction (4.5) in addition to the TVL2-deconvolution (4.6).

tion correction step to correct for individual geometric distortions, illustrated in Figure 4.5. A similarity 2D transformation is used first in (4.5) to correct for rigid motion and uniform in-plane scaling for each slice separately. This is performed twice with different spatial initialisation transformations, i.e. using the identity transformation and the initialisation transformation based on the moments of the skull-masked historical and the reference slice whereby the result with overall lower cost is selected. The prior term in (4.5) is chosen to penalise in-plane scaling and intensity coefficients. Since all slices have been uniformly scaled previously, the prior value for in-plane scaling is set to 1. Similarly, due to the global intensity scaling, the coefficients α_k and β_k are expected to be close to 1 and 0, respectively. Regularisation parameters are found experimentally and described in Section 4.3.2. Afterwards, the full, 2D affine transformation model is chosen for (4.5) so that the historical slice can match the skull as accurately as possible.

After having estimated the positions and geometrical properties of the slices a subsequent image deconvolution step is performed for each slice separately in order to restore each individual slice without mixing neighbouring slice information; this is in contrast to the previous Chapters 2 and 3 to avoid out-of-plane resampling of the sparse, axial data. For this purpose, we rely on a 2D **isotropic total variation (TVL2) deconvolution** step [Beck and Teboulle, 2009b, Rudin et al., 1992]

$$\min_{\mathbf{x}} \left(\frac{1}{2} \|\mathbf{y} - \mathbf{A}(\sigma^2)\mathbf{x}\|_{\ell^2}^2 + \lambda \text{TV}_{\ell^2}(\mathbf{x}) \right) \quad \text{s.t.} \quad \mathbf{x} \geq 0 \quad (4.6)$$

for each individual slice $\mathbf{y} \in \{\mathbf{y}_1, \dots, \mathbf{y}_K\}$ with $\mathbf{y}_k = (y_k(\boldsymbol{\xi}))_{\boldsymbol{\xi} \in \Omega}$ representing the vectorised slice, \mathbf{x} its deblurred version, \mathbf{A} the blurring operator with covariance σ^2 to jointly describe the in-plane blurring of the image acquisition, printing and scanning and $\lambda > 0$ the regularisation parameter. A *matrix-free* implementation is chosen in order to avoid the storage of large matrices [Diamond and Boyd, 2015, Ebner et al., 2017a]. The optimisation problem (4.6) is solved via Alternating Direction Method of Multipliers (ADMM) described in [Boyd et al., 2011]. The implemented, scaled, explicit form of ADMM iterations reads

$$\mathbf{x}^{i+1} := \arg \min_{\mathbf{x} \geq 0} \left(\frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\ell^2}^2 + \frac{\rho}{2} \|\nabla \mathbf{x} - \mathbf{v}^i + \mathbf{w}^i\|_{\ell^2}^2 \right) \quad (4.7)$$

$$\mathbf{v}^{i+1} := \mathbf{S}_{\lambda/\rho}(\nabla \mathbf{x}^{i+1} + \mathbf{w}^i) \quad (4.8)$$

$$\mathbf{w}^{i+1} := \mathbf{w}^i + (\nabla \mathbf{x}^{i+1} - \mathbf{v}^{i+1}) \quad (4.9)$$

with the auxiliary variable \mathbf{v} , the scaled dual variable \mathbf{w} , the Lagrange multiplier ρ and the vectorial soft threshold operator $\mathbf{S}_{\lambda/\rho}$ [Boyd et al., 2011]. Given that the standard Lawson & Hanson algorithm [Lawson and Hanson, 1974] cannot cope with large-scale non-negative least-squares problems several specialised methods have been proposed in the literature to solve minimisation problems like (4.7) such as presented in [Becker and Fadili, 2012, Kim et al., 2010a, Kim et al., 2013]. In this chapter, we used the L-BFGS-B v3.0 solver [Byrd et al., 1995, Morales and Nocedal, 2011] which, although not specifically designed for non-negative least squares, generally shows good performance for such problems and, especially for large-scale problems, regularly outperforms other modern methods [Kim et al., 2010a].

4.3. Data, Evaluation Methodology and Results

4.3.1. Data

A cohort of people recruited soon after a CIS was first assessed at the National Hospital for Neurology and Neurosurgery, Queen Square, London, between 1984-1987 and followed up at regular time points until present [Miller et al., 1988, Miller et al.,

1989, Morrissey et al., 1993, O’Riordan et al., 1998, Brex et al., 2002, Fisniku et al., 2008]; a 30-year longitudinal follow-up, clinical study is currently underway including more than 100 image acquisitions captured on historical films. The preserved MR films were scanned using the Vidar DiagnosticPRO Advantage film digitizer, processed with the Hipax Diagnostic Workstation medical image viewer software and exported to DICOM files. For the current study, a subset of 20 MR film sets (18 acquisitions at 5-year follow-up acquired at 0.5 T and two at 10-year follow-up acquired at 1.5 T) was available where both the original MR films and the stacked, digital scans of the exact same 5 mm thick PD-like slice acquisitions were available. This is a rare situation and allows to validate the volumetric reconstruction pipeline against ground-truth data. In 18 out of those 20 subjects, the acquisitions were captured on films where two consecutive slices showed overlapping skull and brain structures similar to Figure 4.1a and Figure 4.1d. To recover the spatial correspondences for each slice a later PD-scan of the same subject was used as the reference which is typically available in longitudinal studies. The reference scan was acquired as stacks of 5 mm thick slices and, depending on the subject-specific follow-ups and availability, either was a 10-, 14- or 20-year time point after the baseline scan of the same subject. This reference was also used to correct for existing left-right flipping of the brain we encountered in the scans.

4.3.2. Parametrisation of Volumetric Reconstruction Pipeline

The entire, regularised volumetric reconstruction framework was implemented in PYTHON while taking advantage of ITK for the individual registration steps. The joint regularised registration problem (4.5) was implemented via the `least_squares` algorithm of SCIPY where the exact Jacobian was provided for both accelerated and more accurate computational results. The framework is made open source and available on github⁴.

The semi-automatic slice extraction tool stores the naively stacked 3D data as a NIFTI image for further processing. The rigid in-plane registration step with inter-

⁴<https://github.com/gift-surg/VolumetricReconstructionFromPrintedFilms>

slice regularisation using slice neighbour-terms (4.3) was initialised based on the slice moments and used least-squares differences as similarity metric whereby 10 iterations were performed in the `least_squares` algorithm. By considering this stack of neighbour-aligned slices as a 3D volume BET was applied to extract its brain mask. The skull mask was then defined as its negated mask followed by a subsequent dilation step to account for geometric discontinuities across slices. The skull mask for the more recent 3D reference image was obtained analogously but without the dilation step. Constrained by the skull masks, the in-plane 3D similarity registration step was performed using cross-correlation as the similarity measure, linear interpolation resampling, regular step gradient descent optimiser with physical shift scales estimation and a three level multi-resolution framework which was initialised by a previously performed rigid registration based on the respective 3D image moments.

The global intensity correction was performed as described in Section 4.2.4 by using the global 20 %-intensity percentile for the lower threshold. This threshold was found experimentally by visual comparisons.

The motion correction method with inter-slice regularisation and reference image information transfer in combination with the affine intensity correction model (4.5) described in Section 4.2.5 was implemented via the `least_squares` optimiser. The prior term \mathcal{P} was set up to regularize the in-plane scaling and the affine intensity correction parameters only so as to extend the inter-slice regularisation of the motion correction framework. Due to the global scaling and intensity correction performed during the initialisation steps, the associated prior values are set to 1 for the in-plane scaling and $\alpha_{k0} = 0$ and $\beta_{k0} = 1$ for the intensity correction parameters, respectively. By using the solver-specific `soft_11` for the respective λ -weighted residuals in (4.5), the applied smooth ℓ^1 -approximation corresponds to $\varrho_{1/\lambda}(e) = \sqrt{1/\lambda^2 + e^2} - 1/\lambda$ for $\lambda \in \{\lambda_{\mathcal{R}}, \lambda_{\mathcal{N}}, \lambda_{\mathcal{P}}\}$. The weights and the regularisation parameter were found experimentally and set to $\lambda_{\mathcal{N}} = 1$, $\lambda_{\mathcal{R}} = 10$ and $\lambda_{\mathcal{P}} = 10^6$ in (4.5) for the 2D similarity registration step, respectively, whereby 10 iterations were performed. For the subsequent 2D affine registration, the regularisation parameter was reduced to $\lambda_{\mathcal{P}} = 10^3$ and 20 iterations were performed which was sufficient to achieve overall

convergence. During experiments we found that omitting the inter-slice regularisation term in (4.5) can lead to severe misregistrations during motion correction. We, therefore, conclude that the proposed motion correction framework based on inter-slice regularisation is key in order to reliably achieve volumetric reconstructions of high anatomical accuracy. Associated comparisons are provided at the end of Section 4.3.4. For the TV reconstruction step (4.6), the regularisation parameter $\lambda = 5$, the Lagrange multiplier $\rho = 0.5$ and the covariance $\sigma^2 = 0.25$ for the blurring operator \mathbf{A} were found via L-curve studies. The first-order Tikhonov problem (4.7) in the corresponding TVL2 deconvolution step with its positivity constraints was solved via the L-BFGS-B algorithm of SciPY to iteratively solve for the unique minimizer whereby 10 ADMM iterations were performed.

4.3.3. Evaluation Methodology

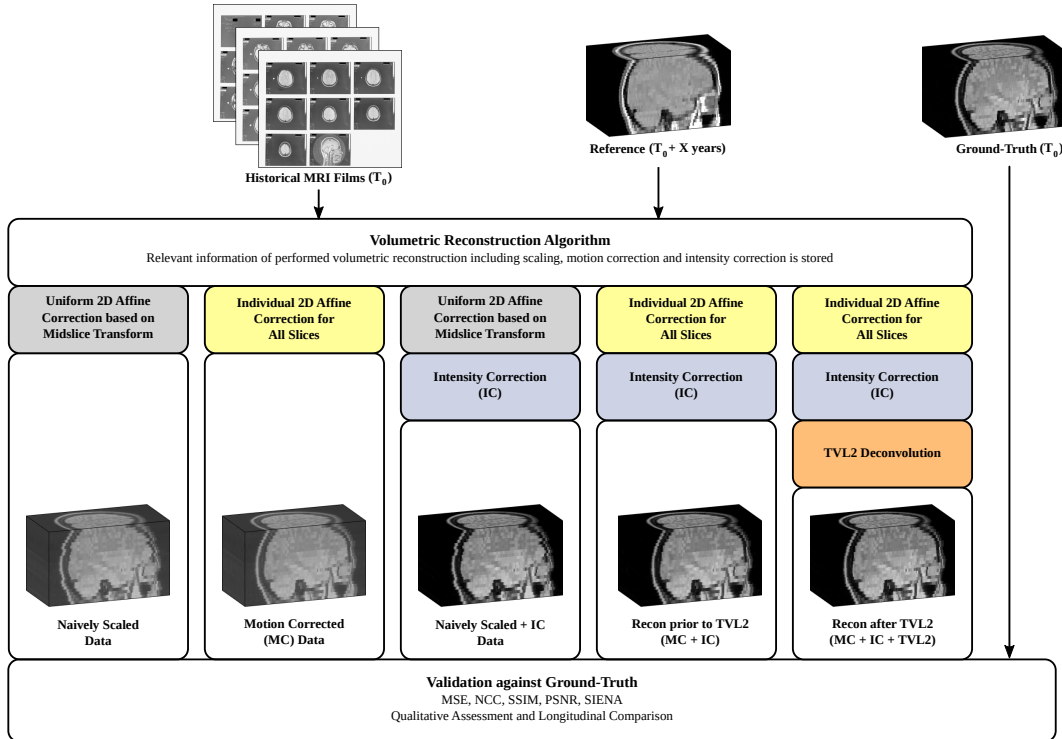


Figure 4.6.: Summary of the pipeline used to validate the volumetric reconstruction framework. Each image shows a typical volumetric output obtained after the respective steps of processing.

Several similarity measures like mean squared error (MSE), peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM)[Wang et al., 2004] and normalised cross-correlation (NCC), were used to assess the similarity between ground-truth and different intermediate results until the final volumetric reconstruction with the TVL2 step. We considered the reconstruction prior to the TVL2 step (full 2D affine correction including intensity correction but no TVL2 step), the motion corrected (MC) data (full 2D affine correction but no intensity correction), the naively scaled data (naively stacked data scaled based on the final 2D affine transformation belonging to the mid-slice of the stack but no intensity correction), the naively scaled and intensity corrected (IC) data (same intensity correction applied to the naively scaled data), and the reference used for motion and intensity correction. A visual summary of the used validation pipeline including the respective short-hands is shown in Figure 4.6. The similarity measures were only evaluated at the masked brain region obtained via BET [Smith, 2002] applied on the ground-truth. The required alignment of stacks prior to the evaluation was obtained by using the rigid registration algorithm `reg_aladin` within NIFTYREG⁵ which is based on block-matching [Modat et al., 2014].

In addition, we evaluated the absolute mean surface motion in linear voxel units of the reconstruction after TVL2 step and the naively scaled and intensity corrected data compared to the ground-truth which reflects the sum of all edge motions between two segmentations [Smith et al., 2002]. This measure was computed via Structural Image Evaluation, using Normalization, of Atrophy (SIENA) [Smith et al., 2002], where we only measured in-plane edge motion because of the missing top brain on the historical data, as visible in Figure 4.2.

Following this, a subjective quality assessment in a clinical context was performed where two blinded neurologists assessed the naively scaled and intensity corrected data, the reconstruction prior to TVL2 and the reconstruction after TVL2 step side-by-side and in comparison with the naively scaled data, the ground-truth data and the reference image used for motion and intensity correction. After performing a

⁵<https://sourceforge.net/projects/niftyreg>

contrast auto-adjustment in the image viewer for more comparable visualisation, scores were given for:

1. Clinical usefulness ranking based on lesions' conspicuity and geometric plausibility/skull continuity essential for volumetric measurements in addition to a final score on overall preference.
2. Ground-truth comparison in terms of interpretability based on image quality and anatomical similarity.

4.3.4. Results

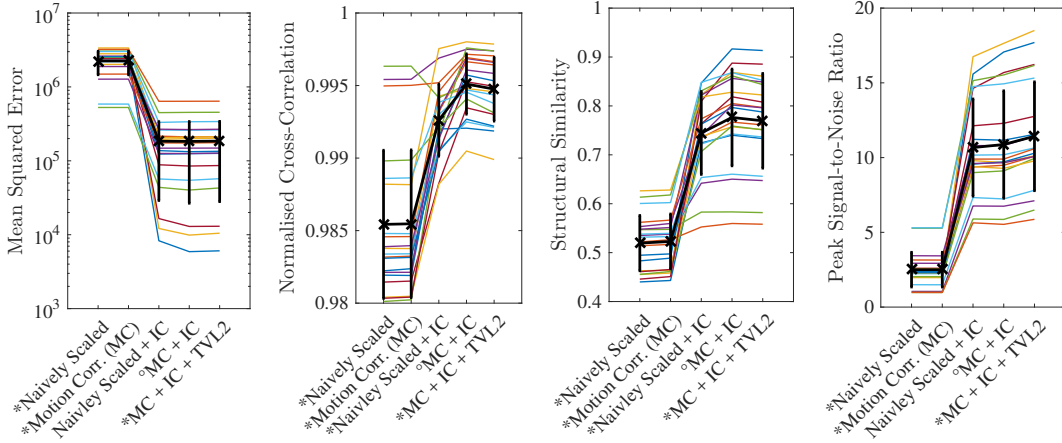


Figure 4.7.: Similarity measures evaluated at the ground-truth brain for each subject separately. Each of the 20 subjects is assigned a different curve. The bold black curve shows the error bar given by mean and standard deviation. A * indicates that the reconstructions are statistically significantly different from the reconstruction prior to TVL2 (MC + IC, marked with a °) based on a paired t -test ($p < 0.05$) and Bonferroni-correction.

Figure 4.7 illustrates that the biggest improvement in measured similarity is achieved by the intensity correction step. Importantly, however, motion correction is shown to significantly increase the similarity to the ground-truth. A further, significant improvement in PSNR is achieved by the TVL2 deconvolution step at the expense of measured similarity with the ground-truth. This can be explained by the fact that the considered ground-truth stack is affected by blurring and noise due to the acquisition performed decades ago. Therefore, deblurring can counteract the PSF during

Table 4.1.: Summary of similarity measures evaluated at the ground-truth brain stated as mean and standard deviation for all 20 subjects. The MSE was omitted in favour of less absolute intensity value sensitive measures. The symbol \neq^* indicates a statistically significant difference between the left and right hand-side with respect to the statistics shown in the sub-index based on a paired t -test ($p < 0.05$) and Bonferroni correction.

	NCC	SSIM	PSNR	Notes
a) Naively Scaled	0.985 ± 0.005	0.519 ± 0.056	2.505 ± 1.159	
b) Motion Corr. (MC)	0.985 ± 0.005	0.523 ± 0.056	2.503 ± 1.160	$b \neq_{NCC/SSIM}^* a$
c) Naively Scaled + IC	0.993 ± 0.002	0.745 ± 0.085	10.671 ± 3.250	$c \neq_{NCC/SSIM/PSNR}^* a, b$
d) MC + IC	0.995 ± 0.002	0.776 ± 0.099	10.876 ± 3.589	$d \neq_{NCC/SSIM}^* a-c$ & $d \neq_{PSNR}^* a, b$
e) MC + IC + TVL2	0.995 ± 0.002	0.770 ± 0.097	11.426 ± 3.639	$e \neq_{NCC/SSIM/PSNR}^* a-d$

acquisition time and have a positive impact on the image quality [Beck and Teboulle, 2009b, Rudin et al., 1992, Buades et al., 2005]. The evaluation in Table 4.1 allows a more detailed assessment of the contribution of each individual step as outlined in Figure 4.6. It clearly shows that motion correction applied on both intensity and non-intensity corrected data leads to significant improvements in measured image similarities. It also illustrates the high intensity contrast dependency of all involved measures which explains the visually striking impact of the performed non-linear intensity correction observed in Figure 4.7. However, by considering the same starting point of either non-intensity or intensity corrected data, the significant similarity improvements by the performed motion correction underline its importance and effectiveness to obtain high-quality volumetric reconstructions.

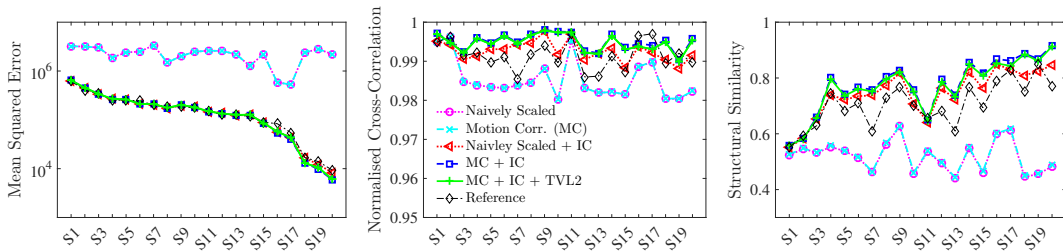


Figure 4.8.: Similarity measures evaluated at the ground-truth brain for each individual subject. The subjects on the horizontal axis are ranked according to the MSE of the reference image, shown for comparison purposes.

Figure 4.8 provides a per-subject comparison and shows the impact of each performed step in the volumetric reconstruction pipeline for each individual case. The

high figures in NCC and SSIM of the naively scaled and intensity corrected data reveal that the semi-automatic slice-extraction is able to achieve an overall high accuracy of initial slice alignment which is further, significantly, improved by the volumetric reconstruction pipeline. The MSE suggests that the performed intensity correction is very effective for all subjects and yields slice intensities similar to their respective references. The figures of NCC and SSIM confirm that the motion correction had a significant impact and led to substantial improvements in image similarity for almost all subjects. Both NCC and SSIM also show that our volumetric reconstructions achieve higher similarity than the reference images illustrating the morphological changes the brain has undergone over time and differences owed to the different contrast.

SIENA measured the absolute mean surface motion between ground-truth and naively scaled image as 0.80 ± 0.20 linear voxel units with 95 %-confidence interval (CI) from 0.71 to 0.88. The reconstruction after TVL2 step achieved a mean figure of 0.61 ± 0.13 with 95 %-CI from 0.56 to 0.67 which corresponds to a significant improvement based on paired t -test ($p < 0.05$) of about 25 %. Therefore, detected edge-motion was significantly reduced which underlines the capability of the performed motion correction framework to significantly increase the accuracy of the obtained volumetric reconstructions.

Table 4.2.: Summary of blinded, clinical evaluation averaged over all 20 subjects. Lesions conspicuity and geometric plausibility/skull continuity subjectively rank preferred reconstruction from 1 (least preferred) to 3 (most preferred). Preferred image score indicates the number of times the respective reconstruction was the preferred choice. Ties were allowed for the geometric plausibility and preference ranking in case images were visually indistinguishable. The anatomical similarity to the ground-truth is rated 0 (distinguishable) or 1 (not distinguishable). Image quality similarity to ground-truth scores are 0 (worse, but interpretable), 1 (same as ground-truth) and 2 (improved interpretation).

	Clinical Usefulness Ranking			Ground-Truth Comp.	
	Lesions' Conspicuity	Geometric Plausibility	Pref. Image	Anatomical Similarity	Image Quality
Naively Scaled + IC	1.45 ± 0.69	2.00 ± 0.00	0	0.10 ± 0.31	0.70 ± 0.73
MC + IC	2.40 ± 0.50	3.00 ± 0.00	8	0.80 ± 0.41	1.35 ± 0.59
MC + IC + TVL2	2.85 ± 0.37	3.00 ± 0.00	15	0.95 ± 0.22	1.50 ± 0.51

To further investigate the impact of the performed motion correction and denoising steps we analysed the reconstruction quality of the associated volumes and in direct comparison with the naively scaled and intensity corrected data by performing a qualitative assessment by expert neurologists. The neurologist’s evaluation, shown in Table 4.2, indicates that the blinded neurologists had a clear preference for our volumetric reconstructions over the naively scaled data given their higher geometric plausibility and improved lesions’ conspicuity. Adding the TVL2 deconvolution step yields even further improvement. Particularly, the score on geometric plausibility states that the performed motion correction always yielded an improved outcome. In direct comparison, we almost always achieve results which are visually indistinguishable from the ground-truth data. In addition, it was felt that especially the reconstruction after TVL2 step gives rise to improved interpretation; better than the original, non-processed ground-truth data which itself is affected by its PSF-affected physical acquisition from the past.

In Figure 4.9 the naively scaled stack and the reconstruction results before and after the TVL2 step are provided for one of the cases along with the ground-truth data and the reference image used for motion and intensity correction. This example was selected to showcase the result for one of the visually most motion corrupted stacks after the semi-automatic slice extraction step which served as initialisation of the volumetric reconstruction framework. The overlaid brain mask of the ground-truth illustrates the discrepancy of the naively scaled data which becomes almost invisible in the obtained volumetric reconstruction results. Only at the neck, slight inaccuracies of the reconstructions become apparent in the sagittal view which can be explained by the high intensities at this region on the reference image. The bottom row of Figure 4.9 shows a zoomed-in comparison highlighting the high accuracy of the motion correction in combination with the image quality improvements due to intensity correction and deconvolution steps.

Figure 4.10 provides a qualitative comparison of three longitudinal scans used in this study associated with a subject where the baseline scan was acquired in 1986. The visualised subject represents the only available case in this study where both

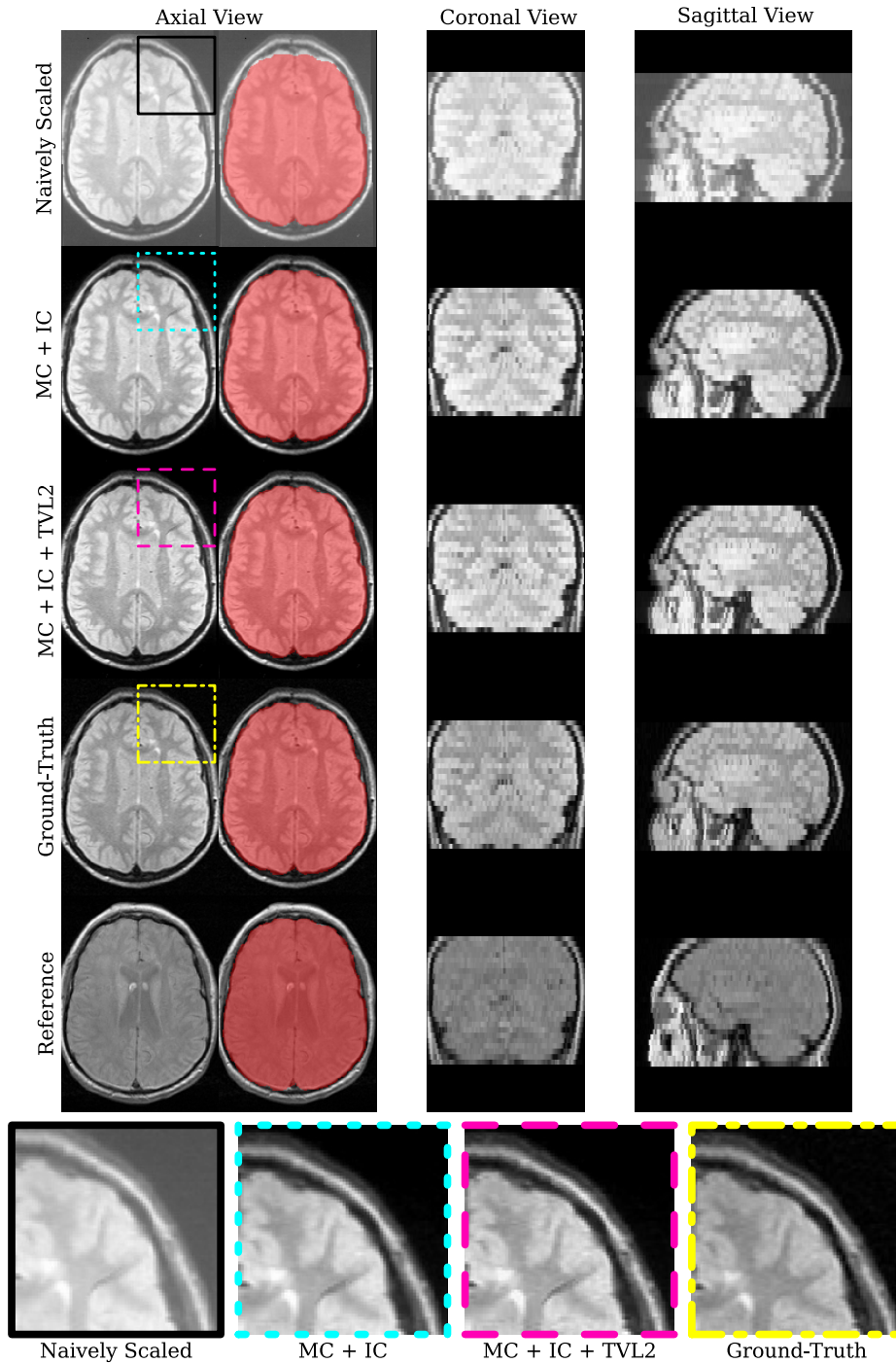


Figure 4.9.: Visual comparison between reconstruction results and original data. The automatically segmented ground-truth brain is shown as a red overlay in each of the images and illustrates the reconstruction accuracy of the obtained volumetric reconstruction from limited FOV data. At the bottom, a zoomed window highlights the improvement and reconstruction quality. There, also a stroke of dirt is visible on the historical data.

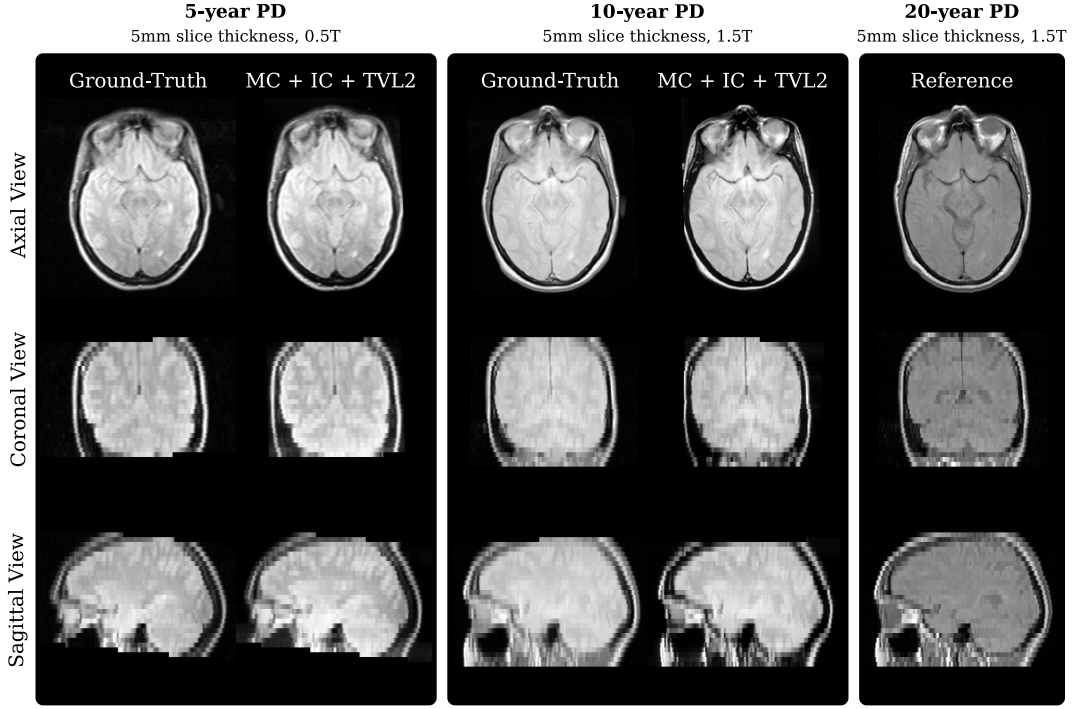


Figure 4.10.: Qualitative comparison of three longitudinal scans used in this study associated with a subject where the baseline scan was acquired in 1986. The comparison shows the 20-year scan used as reference image for the volumetric reconstruction pipeline along with the linearly resampled digital, ground-truth, data, and the obtained volumetric reconstructions from the historical films of the respective 5-year and 10-year follow-up time points. Visually, differences between the reconstructions and the ground-truth are hardly detectable. The measured ground-truth-similarities for the reconstructed 5-year scan are 0.992, 0.787, 10.294, for NCC, SSIM and PSNR, respectively. Similarly, the respective figures for the 10-year scan are 0.993, 0.844, and 16.167. Note that due to overlapping skulls on the historical MR films, see Figure 4.1a, only the part visible on the films could be recovered during the semi-automatic slice extraction step for the 10-year scan.

a 5-year and 10-year digital stack, i.e. a “ground-truth”, were available. One can observe the highly consistent reconstructions obtained by the proposed volumetric reconstruction framework with closely matching contours of skull and brain for both 5- and 10-year follow-up scans.

Using our non-optimised implementation, the typical processing time to restore one single subject from printed MR films was measured to be approximately 1 h 20 min on a single computer. This includes about 2 min to 5 min of user interaction to operate the semi-automatic slice extraction tool. The remaining processes are fully automated whereby the volumetric reconstruction steps, including motion and intensity correction, were measured to take about 45 min and the final TVL2 deconvolution step about 30 min.

The Importance of Inter-slice Regularisation

In this section, we summarise additional experiments to underline the importance of using inter-slice regularisation in the proposed motion correction framework.

For this purpose, we volumetrically reconstructed the historical data described in Section 4.3.1 with the method described in Section 4.2 by using three variants of the cost function (4.5). Using the parametrisation of the volumetric reconstruction pipeline as described in Section 4.3.2, we reconstructed the historical data in three experiments characterized by increasing model complexity by varying the parameters $\lambda_{\mathcal{R}}$, $\lambda_{\mathcal{N}}$ and $\lambda_{\mathcal{P}}$. Specifically, we ran the reconstruction pipeline a) without using regularisation for motion correction, i.e. motion correction is based on the reference similarity term \mathcal{R}_k alone, b) without using inter-slice regularisation, i.e. using the reference similarity term \mathcal{R}_k and the regularisation term \mathcal{P} but no inter-slice similarity term \mathcal{N}_k , and c) using our proposed method. The three experiments were run based on the parameters

- a) $\lambda_{\mathcal{R}} = 10, \lambda_{\mathcal{N}} = 0, \lambda_{\mathcal{P}} = 0$, i.e. no regularisation at all
- b) $\lambda_{\mathcal{R}} = 10, \lambda_{\mathcal{N}} = 0, \lambda_{\mathcal{P}} = 10^3$, i.e. no inter-slice regularisation
- c) $\lambda_{\mathcal{R}} = 10, \lambda_{\mathcal{N}} = 1, \lambda_{\mathcal{P}} = 10^3$, i.e. method as proposed

By using a similar comparison as in Table 4.1, we quantitatively compared the reconstruction quality achieved by all three variants for all 20 subjects to investigate the importance of using inter-slice transform regularisation in (4.5). The results are summarised in Table 4.3 and show that the measured ground-truth similarity is consistently higher in terms of both NCC and SSIM by using the proposed method with inter-slice regularisation and outperforms any of the “reduced” methods a) and b). This can be explained by the fact that the insufficient transform regularisation of the motion correction based on a) or b) can lead to severe misregistrations of individual slices and, subsequently, to volumetric reconstructions of poor anatomical plausibility.

Additionally, we provide a qualitative comparison for one subject to illustrate the insufficiency of the methods a) and b) to reliably achieve high-quality reconstructions in Figure 4.11 which summarises our experience: Using motion correction based on inter-slice regularisation reduces the risk of slice misregistration and generally achieves volumetric reconstructions of high anatomical accuracy.

Table 4.3.: Summary of similarity measures evaluated at the ground-truth brain stated as mean and standard deviation for all 20 subjects. The MSE was omitted in favour of less absolute intensity value sensitive measures. Bold values correspond to the best outcome. A * indicates that the reconstructions are statistically significantly different from the result obtained by the proposed method based on a paired t -test ($p < 0.05$) and Bonferroni-correction.

	NCC	SSIM	PSNR
a) Motion Corr. (MC) without reg.	0.985±0.005	0.521±0.056*	2.505±1.155
b) MC without inter-slice reg.	0.985±0.005	0.522±0.056	2.503±1.159
c) MC with proposed method	0.985±0.005	0.523±0.056	2.503±1.160
a) MC + IC without reg.	0.992±0.006	0.754±0.100*	10.780±3.285
b) MC + IC without inter-slice reg.	0.994±0.002*	0.758±0.093*	10.628±3.245
c) MC + IC with proposed method	0.995±0.002	0.776±0.099	10.876±3.589

4.4. Discussion

In this chapter, we presented, and to the best of our knowledge for the first time, a framework which reconstructs the volumetric stack from printed, historical, limited FOV MR films being acquired decades earlier. The proposed semi-automatic

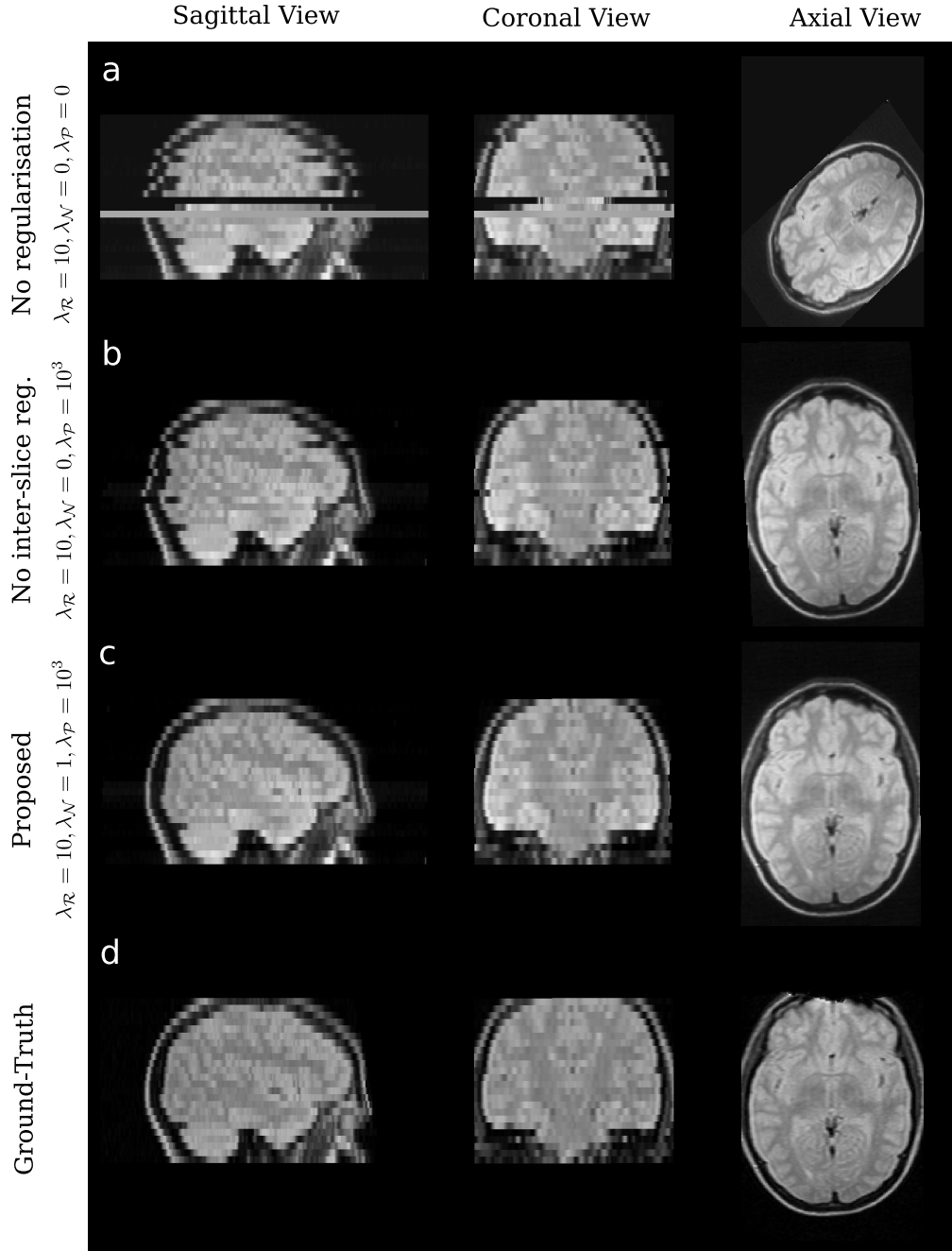


Figure 4.11.: Reconstructed sample to show the importance of inter-slice regularisation for motion correction to obtain high-quality volumetric reconstructions. Severe misregistrations are visible in case motion correction is based on the reference image information alone (a). Adding the regularization prior term \mathcal{P} achieves slight improvements (b). A volumetric reconstruction of high anatomical accuracy is achieved by using the proposed reconstruction pipeline based on the minimization problem (4.5) which additionally relies on inter-slice regularisation (c).

slice-extraction algorithm is capable of dealing with different MR films of many kinds, formats and appearances including films where parts of the brain and skull are overlapping on consecutively printed slices. Its particular design choice ensures the robustness to any acquisition set-up with respect to slice-ordering or single/multi-slice acquisition in the sense that printed slices which capture adjacent anatomy are reliably combined to form a first naively stacked 3D data for further processing regardless of the encountered type of historical MR film. We introduce a joint slice-to-volume registration with inter-slice transformation regularisation and slice intensity correction based on a smooth ℓ^1 approximation as loss functional for a more robust registration framework. We put a particular focus on using methods which are able to restore the original image quality and geometry of the historical scans without introducing additional image artefacts. We emphasise on numerical accuracy and computational efficiency by providing the exact Jacobian for all cost functions and use a matrix-free implementation during the TVL2 deconvolution step. We test our reconstruction results against ground-truth data both quantitatively and in a clinical context and demonstrate the high reconstruction quality and suitability of the proposed framework.

The performed validation proves the used global intensity correction step sufficient to deal with the existing intensity differences and illustrates its substantial contribution to an overall improved similarity to the ground-truth data. It demonstrates that the motion correction algorithm yields a further, statistically significant, improvement towards overall similarity in both the measures of structural similarity and normalised cross-correlation. It underlines that the proposed motion correction is vital to eliminate implausible discontinuities across slices existing after the semi-automatic slice extraction and yields reconstructions with highly plausible brain geometries which accurately reflect the patient-specific anatomy. It illustrates that the volumetric reconstruction framework and its design is capable of robustly reconstructing accurate volumetric reconstructions from historic MR films even when skulls are merged and, consequently, information on the skull is compromised. It shows that the additional TVL2 deconvolution step gives rise to volumetric recon-

structions which are visually almost indistinguishable from the ground-truth data and may even lead to an improved interpretation over the original, digital volumes.

The volumetric reconstruction algorithm with its joint, in-plane 2D affine motion and intensity correction model in addition to the in-plane 2D deconvolution step is designed as a framework to carefully balance the desire of fully recovering the original 3D image anatomy without mixing slice neighbour information or introducing image artefacts. However, this approach can account for axial motion only and may well be insufficient in cases where subject motion occurred during acquisition time or more complex motion is present in the historical MR films. The consistently high reconstruction quality shown in the course of the validation supports the argument that inter-slice subject motion was not an issue for the data in our study. Our proposed pipeline is not designed to reconstruct parts of the brain which are not visible in the original MR films. This includes anatomical information hidden by partial voluming effects due to the thick slice acquisitions or parts of the brain which are entirely missing due to a reduced FOV. Hence, a truncation at the vertex will prevent whole brain volumes from being estimated. Furthermore, the anatomical accuracy of the volumetric reconstruction depends on the estimate of the respective, oblique reference slices obtained by resampling of the reference stack. Thus, a higher resolution of the reference image would provide more scope of accurate registrations. Minor issues we encountered with our method were associated with subjects where either the top or the bottom slices of the historical stack did not have a sufficient overlap with the FOV of the reference image. For such slices only the slice neighbour-term (4.3) contributes to the cost function (4.5). This situation effectively simplifies to an intra-stack in-plane registration problem and tends to align bordering slices so that they fit their proximal neighbour. Consequently, minor inaccuracies for these slices are visible which may slightly differ from the correct anatomical shape.

This volumetric reconstruction pipeline was developed based on the assumption that a reference 3D digital image exists with similar intensity appearance in order to extract both its meta-data and intensity contrast information. However, in cases where no such reference is available the proposed volumetric reconstruction pipeline

could still be applied in various ways. In case a naive digital 3D representation is sufficient for the problem at hand the naively stacked data after semi-automatic slice extraction can be used whereby the imaging meta-data could be updated with manual values. Additional motion correction could be performed by using intra-stack regularisation (4.3) only. However, this is likely to lead to non-physiological slice-alignments like a straight skull delineation. Alternatively, an atlas could be used to apply the entire volumetric reconstruction method as outlined in Section 4.2. Nevertheless, a high reconstruction accuracy depends on the possibility to realign all slices so that they match the patient-specific anatomy as closely as possible. For this step, a reference image which accurately reflects the subject anatomy is key and the higher the slice thickness becomes the more important such a similarity will be for our proposed framework.

Overall, our framework has shown its capability to accurately reconstruct 3D volumes from printed MR films of MS subjects and will help in the robust analysis of a uniquely long-term study spanning 30 years of MRI scans of people followed up after a CIS. This study includes more than 100 subject scans captured on historical films which can be reconstructed with our proposed technique for further image processing and analysis. However, applications of the proposed method are not confined to CIS and MS studies, and it may prove useful for the longitudinal assessment of lesions and anatomical structures in a variety of other conditions that affect the brain. Our framework may also be useful in current clinical practice, where not uncommonly, patients have had MR imaging performed previously on different scanners, and the only format available is in film format. The volumetric reconstruction of these images would allow the digital storage of the data, and also a longitudinal comparison with an up-to-date scan. Moreover, despite being tested on only PD-like images where ground-truth data was available, the proposed framework may be used for other MR image contrasts as well.

CHAPTER 5.

Forward-Backward Splitting in Deformable Image Registration: A Demons Approach

Table of Contents

5.1. Introduction	189
5.2. Forward-Backward Splitting for Image Registration	191
5.2.1. Forward-Backward Splitting Methods	191
5.2.2. Forward Step in Image Registration	193
5.2.3. Proximal Step in Image Registration	193
5.3. Experiments	195
5.3.1. Circle to C	195
5.3.2. 3D Anatomical MRI	196
5.4. Discussion	197

Foreword. This chapter is adapted and contains content from the work published in

Ebner, M., Modat, M., Ferraris, S., Ourselin, S., and Vercauteren, T. (2018b). Forward-backward splitting in deformable image registration: A demons approach. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1065–1069, Washington, D.C. IEEE.

In the previous chapters, volumetric reconstruction frameworks were presented that critically relied on accurate registration/motion correction. All associated numerical approximations to estimate the slice transformation parameters in each step relied on traditional solvers like conjugate gradient methods, whereby the problem of convergence and optimality of these solvers was not investigated. This chapter provides an opportunity to take a deeper look at the problem of registration and to potentially improve the efficacy and reliability of these steps.

My specific contributions presented in this chapter are:

- Introduction of recently proposed Forward-Backward Splitting (FBS) methods to deformable image registration.
- Highlighting that FBS is guaranteed to converge to a critical point for cost functions of the form $f + g$ consisting of a smooth (possibly non-convex) function f and a convex (possibly non-smooth) function g . In the context of (non-linear) image registration, f will typically correspond to a (differentiable) similarity measure and g to a (convex) regularizer.
- Showcasing the advantage of FBS to efficiently solve for various kinds of cost functions vital for medical image registration by performing two simple iterative steps – a *forward* and a *backward* step.
- Illustration that Tikhonov regularization breaks down to simple B-Spline filtering in the backward step.
- Demonstration of the versatility of FBS by encoding spatial transformation as displacement fields or free-form B-Spline deformations.
- Comparison of two FBS variants, namely FISTA [Goldstein et al., 2014] and iPiano [Ochs et al., 2014], against the classical demons algorithm, the recently proposed inertial demons algorithm [Santos-Ribeiro et al., 2016] and the conjugate gradient method within NIFTYREG [Modat et al., 2010].
- Numerical experiments performed on both synthetic and clinical data show the advantage of FBS in image registration in terms of both convergence and accuracy.

5.1. Introduction

Efficient non-linear image registration implementations are essential in the fields of medical and biomedical imaging to allow for both timely and accurate modern image analysis. In its general formulation, image registration aims at finding a (non-linear) registration transform T which best aligns a fixed with a moving image, denoted by F and M , respectively. This is typically framed as the optimization of a global energy

$$E(T) := \text{Sim}(F, M \circ T) + \text{Reg}(T) \quad (5.1)$$

consisting of a similarity measure Sim and an additional regularization term Reg to better constrain the otherwise ill-posed problem. Given the difficulty of solving the (in most applications of interest) non-convex problem (5.1) directly, in the *demons* algorithm, the relaxed formulation,

$$\tilde{E}(S, T) := \text{Sim}(F, M \circ S) + \lambda \|S - T\|^2 + \text{Reg}(T) \quad (5.2)$$

with $\lambda > 0$, is optimized instead by solving iteratively for the spatial transformations S and T [Thirion, 1998, Pennec et al., 1999, Cachier et al., 2003, Vercauteren et al., 2007, Vercauteren et al., 2009, Santos-Ribeiro et al., 2016]. In the first step of this demons approach, T is being fixed and the optimization of $\text{Sim}(F, M \circ S) + \lambda \|S - T\|^2$ with respect to S is approximated by a gradient descent step. Then, the updated S is fixed and the optimization of $\lambda \|S - T\|^2 + \text{Reg}(T)$ with respect to T is performed with a simple Gaussian filter applied to S . Both steps are then repeated until convergence. Despite its simplicity, it has been shown to be a very powerful approach in practice. Much work has been invested to further extend the demons approach. For instance, in [Vercauteren et al., 2009] the incorporation of diffeomorphic transformations was suggested to enforce preservation of topology and, more recently, it was suggested to incorporate an additional inertial term to improve overall convergence speed and accuracy [Santos-Ribeiro et al., 2016]. The diffusion-like Gaussian regularization

was extended to bilateral filtering to allow for deformation discontinuities in [Papież et al., 2014] and a duality-based Total Variation (TV) approach for optical flow was proposed in [Pock et al., 2007] to solve a convex approximation of form (5.2). Thus, a wide range of algorithms typically solve a *relaxed* formulation which is not guaranteed to obtain an optimal solution of the original formulation (5.1). In contrast, gradient descent and conjugate gradient methods have been applied to directly solve the original problem (5.1) for a variety of cost functions and regularizers, e.g. [Modat et al., 2010]. However, some form of relaxation is typically employed in case the regularizing term is non-differentiable as in the setting of TV regularization.

In this chapter, we want to highlight a numerical framework which indeed is able to solve the deformable registration problem (5.1) in its *original* form. As an active field of research, Forward-Backward Splitting (FBS) methods have been developed to solve convex and, more recently, non-convex problems of the form

$$\min_{\mathbf{u} \in \mathbb{R}^N} (f(\mathbf{u}) + g(\mathbf{u})) \quad (5.3)$$

whereby the composite objective function consists of a smooth (possibly non-convex) function $f : \mathbb{R}^N \rightarrow \mathbb{R}$ and a convex (possibly non-smooth) function $g : \mathbb{R}^N \rightarrow \mathbb{R}$ [Goldstein et al., 2014, Ochs et al., 2014]. In case the associated proximal operator of g , i.e.

$$\text{prox}_g(\mathbf{u}, \tau) := \arg \min_{\mathbf{v} \in \mathbb{R}^N} (\|\mathbf{u} - \mathbf{v}\|^2 + 2\tau g(\mathbf{v})), \quad (5.4)$$

with $\tau > 0$ is easy to compute, FBS methods break down the iterative solution of (5.3) into two simple, iterative steps: a *forward* gradient descent step on f and a so-called proximal, *backward* gradient descent step. Their advantage, however, lies in their sound mathematical basis, proof of algorithmic convergence to a critical point, simplicity to use and their richness of possible functions f and g .

In this chapter, (i) we propose using a guaranteed-to-converge FBS framework to solve directly for non-linear registration problems of the form (5.1); (ii) we demonstrate their advantage to efficiently implement various kinds of regularizers vital for

medical image registration; (iii) we specifically illustrate that Tikhonov regularization breaks down to simple B-Spline filtering in the proximal step; (iv) we showcase both the fast iterative shrinkage/thresholding algorithm (FISTA) [Goldstein et al., 2014] and the recently proposed inertial proximal algorithm for non-convex optimization (iPiano) [Ochs et al., 2014] as FBS instances; (v) we illustrate the versatility of FBS by relying on a combination of displacement fields, free-form B-Spline transformation, sum of squared differences (SSD) and normalized mutual information (NMI) in conjunction with second-order Tikhonov regularization; and (vi) we evaluate the performance of both FISTA and iPiano against the classical demons, the recently proposed inertial demons [Santos-Ribeiro et al., 2016] and the conjugate gradient method within NIFTYREG [Modat et al., 2010].

5.2. Forward-Backward Splitting for Image Registration

5.2.1. Forward-Backward Splitting Methods

Given a differentiable (possibly non-convex) function f and a convex (possibly non-differentiable) function g , problems of the form (5.3) can efficiently be solved by FBS methods¹ in cases where the proximal operator (5.4) of g can be evaluated easily. In case of a convex f , the basic FBS algorithm

```

while not converged do
  |  $\tilde{\mathbf{u}}_{k+1} := \mathbf{u}_k - \tau_k \nabla f(\mathbf{u}_k)$  (Forward step)
  |  $\mathbf{u}_{k+1} := \text{prox}_g(\tilde{\mathbf{u}}_{k+1}, \tau_k)$  (Proximal step)
end

```

guarantees the convergence of \mathbf{u}_k to a critical point $\hat{\mathbf{u}}$ of (5.3) for $k \rightarrow \infty$ for an appropriate step size $0 < \tau_k < 2/L(\nabla f)$, which depends on the Lipschitz constant of the gradient on f only [Goldstein et al., 2014, Ochs et al., 2014]. The recently proposed FBS-variant called “inertial proximal algorithm for non-convex optimization” (iPiano) extends this statement even to *non-convex* functions f [Ochs et al., 2014]. In practice, however, it can be very challenging to estimate the Lipschitz constant

¹Also commonly referred to as *Proximal-Gradient* methods

$L(\nabla f)$ beforehand. By using backtracking, the Lipschitz constant and, hence, the step sizes τ_k can be estimated automatically. Several variants of FBS have been proposed to speed up the convergence of potentially slow FBS. For the numerical experiments we will use both FISTA [Beck and Teboulle, 2009a] and iPiano [Ochs et al., 2014] with their basic iterations shown in Algs. 5.1 and 5.2 and their extended versions in Algs. 5.3 and 5.4, respectively.

Algorithm 5.1: FISTA algorithm [Beck and Teboulle, 2009a]

$\tilde{\mathbf{v}}_{k+1} := \mathbf{u}_k - \tau_k \nabla f(\mathbf{u}_k)$ (Forward step)
 $\mathbf{v}_{k+1} := \text{prox}_g(\tilde{\mathbf{v}}_{k+1}, \tau_k)$ (Proximal step)
 $\alpha_{k+1} := (1 + \sqrt{1 + 4\alpha_k^2})/2$ (Acceleration parameter)
 $\mathbf{u}_{k+1} := \mathbf{v}_{k+1} + \frac{\alpha_k - 1}{\alpha_{k+1}}(\mathbf{v}_{k+1} - \mathbf{v}_k)$ (Prediction step)

Algorithm 5.2: Generic iPiano algorithm [Ochs et al., 2014]

$\tilde{\mathbf{u}}_{k+1} := \mathbf{u}_k - \tau_k \nabla f(\mathbf{u}_k) + \beta(\mathbf{u}_k - \mathbf{u}_{k-1})$ (Forward step with inertia)
 $\mathbf{u}_{k+1} := \text{prox}_g(\tilde{\mathbf{u}}_{k+1}, \tau_k)$ (Proximal step)

FISTA is based on a predefined sequence of acceleration parameters and is characterized by its lack of tuning parameter in addition to its good worst-case performance. The correct step size can be estimated even without knowing the Lipschitz constant by using backtracking line search. Using non-monotone backtracking has the advantage of not discarding iterates with a higher objective value which might still be closer to the minimizer. This is especially useful in case of poorly conditioned problems and alleviates the computational burden since non-monotone line search conditions are less likely to be violated. Generally, FISTA also performs well on non-convex problems although global convergence is not guaranteed. Building on FISTA, the recent iPiano algorithm has been specifically designed for non-convex problems. Its rigorous mathematical analysis ensures favorable properties and guarantees global convergence also for non-convex f under very weak restrictions. As opposed to FISTA, additional tuning parameters are required. This includes the inertial weight $\beta \in [0, 1)$ and the parameters $\eta, c > 1$ which adaptively tune the step size τ_k during run time to achieve fast convergence.

In the context of (non-linear) image registration (5.1) the framework of FBS is ideal for dealing with a wide range of composite objective functions. Here, f will typically correspond to the (differentiable) similarity measure Sim and g to the (convex) regularizer Reg .

5.2.2. Forward Step in Image Registration

To illustrate the use of forward step in FBS, we focus on SSD as the similarity measure and displacement fields as the spatial transformation model. Nonetheless, we highlight and demonstrate in Section 5.3 that FBS allows for both parametric, e.g. free-form B-Spline, and non-parametric spatial transformations. The forward step is based on the gradient of $f(\mathbf{u}) = \text{Sim}(\mathbf{u}) = \text{SSD}(\mathbf{u}) = \frac{1}{2}\|F - M \circ \mathbf{u}\|^2$ with vector field \mathbf{u} . Explicitly written, this reads

$$\nabla_{\mathbf{u}}f(\mathbf{u}(\mathbf{x})) = (M \circ \mathbf{u}(\mathbf{x}) - F(\mathbf{x})) J_{\mathbf{u}(\mathbf{x})} \quad (5.5)$$

with $J_{\mathbf{u}(\mathbf{x})} = \nabla_{\mathbf{u}}M(\mathbf{u}(\mathbf{x}))$ or $J_{\mathbf{u}(\mathbf{x})} = \frac{1}{2}(\nabla_{\mathbf{u}}M(\mathbf{u}(\mathbf{x})) + \nabla F(\mathbf{x}))$ in case symmetric forces are applied at a point \mathbf{x} [Vercauteren et al., 2009]. Moreover, any other differentiable, parametric or non-parametric spatial transform can be incorporated without restriction. This includes the exponential map vital for image registration on the Lie algebra of diffeomorphisms [Ashburner, 2007, Vercauteren et al., 2008, Vercauteren et al., 2009].

5.2.3. Proximal Step in Image Registration

An explicit or easy-to-solve proximal map prox_g is critical for the efficient computation of the proximal step in FBS. We discuss important examples for image registration.

It is well known that Gaussian blurring G_σ with standard deviation $\sigma > 0$ used in the demons algorithm corresponds to applying **diffusion-like regularization**. Based on the relationship between Gaussian smoothing and solving the heat equation established in [Nielsen et al., 1996] we infer the closed-form solution for the proximal

operator for diffusion-like regularization, i.e.

$$G_\sigma * \mathbf{u} = \arg \min_{\mathbf{v}} \left(\|\mathbf{u} - \mathbf{v}\|^2 + \sum_{i,j=1}^D \sum_{k=1}^{\infty} \frac{\sigma^{2k}}{2^k k!} \left\| \frac{\partial^k}{\partial x_i^k} v_j \right\|^2 \right) = \text{prox}_{g_\sigma^\tau}(\mathbf{u}, \tau) \quad (5.6)$$

with $g_\tau^\sigma(\mathbf{v}) = \frac{1}{2\tau} \sum_{i,j=1}^D \sum_{k=1}^{\infty} \frac{\sigma^{2k}}{2^k k!} \left\| \frac{\partial^k}{\partial x_i^k} v_j \right\|^2$ whereby $\mathbf{v} = (v_j)_{j=1}^D : \mathbb{R}^D \rightarrow \mathbb{R}^D$ and D denoting the dimension of space. Hence, Gaussian filtering corresponds to using the function g_τ^σ as regularizer and using a step size $\tau_k = \tau$ for its proximal operator. With a view to FBS where the same step size is used for both the forward and proximal step, this means that adjusting the step size τ_k at FBS iterations implicitly corresponds to an adaptive scaling of the objective function according to $f + g_{\tau_k}^\sigma = f + \frac{1}{\tau_k} g_1^\sigma$ in (5.3). However, proofs of convergence are based on a fixed g and, hence, the strong mathematical statement of guaranteed algorithmic convergence of FBS to a critical point would be lost. Note that (5.6) corresponds to the second step in the classical demons approach too.

Another class of important closed-form proximal maps are related to the application of **Tikhonov regularization**. Based on the insights in [Unser et al., 1993], we explicitly state the proximal operator for r -th order Tikhonov regularization, i.e. $g_r^\lambda(\mathbf{v}) = \frac{\lambda}{2} \sum_{i,j=1}^D \left\| \frac{\partial^r}{\partial x_i^r} v_j \right\|^2$ with regularization parameter $\lambda > 0$, as

$$B_{2r-1,\tau\lambda}(\mathbf{v}) = \arg \min_{\mathbf{v}} \left(\|\mathbf{u} - \mathbf{v}\|^2 + \tau\lambda \sum_{i,j=1}^D \left\| \frac{\partial^r}{\partial x_i^r} v_j \right\|^2 \right) = \text{prox}_{g_r^\lambda}(\mathbf{v}, \tau) \quad (5.7)$$

whereby $B_{2r-1,\tau\lambda}$ denotes the B-Spline smoothing filter of order $2r - 1$ and smoothing parameter $\tau\lambda$. Varying step sizes τ_k for adaptive FBS schemes can easily be incorporated by scaling the smoothing parameter for the filtering without changing the original objective function, in contrast to Gaussian smoothing. Importantly, second-order Tikhonov regularization represents a cubic spline smoothing for the proximal step which can be efficiently implemented using $15D$ operations per voxel by recursive infinite impulse response (IIR) filters [Unser et al., 1993].

Moreover, more complicated regularizers can easily be wrapped into the FBS

framework. For example, **isotropic TV regularization** leads to a proximal step of

$$\text{prox}_g(\mathbf{u}, \tau_k) = \arg \min_{\mathbf{v}} \left(\|\mathbf{u} - \mathbf{v}\|^2 + 2 \tau_k \lambda \text{TV}_{\ell^2}(\mathbf{v}) \right) \quad (5.8)$$

which corresponds to a TV denoising step. By using its dual formulation, isotropic TV regularization can be solved via a *nested* FBS scheme in which a FBS is applied also for the minimization in the proximal step (5.8) [Goldstein et al., 2014].

5.3. Experiments

5.3.1. Circle to C

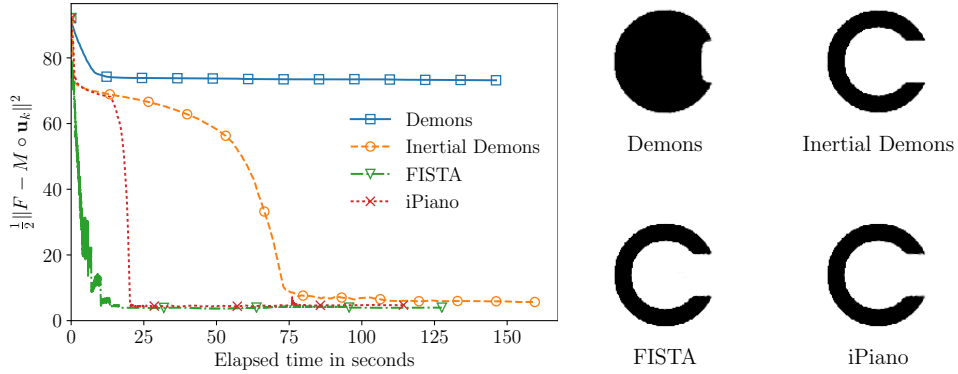


Figure 5.1.: Convergence comparison of solvers for the non-linear registration of the classical “Circle to C” experiment with markers indicating every 500 iterations.

To examine the ability for the proposed FBS framework to perform large deformations, we chose the classical “Circle to C” experiment. We compare both FISTA and iPiano against the additive demons and the recently proposed inertial demons [Santos-Ribeiro et al., 2016]. Based on the similarity measure $\text{SSD}(\mathbf{u}) = \frac{1}{2} \|F - M \circ \mathbf{u}\|^2 = f(\mathbf{u})$, we parameterized the inertial demons as proposed in [Santos-Ribeiro et al., 2016] and implemented the Efficient Second-order Minimization (ESM) [Vercauteren et al., 2009] based gradient in (5.5), set the inertial weight $\alpha = 0.9$ and the maximal step to 0.5 voxels for the demons algorithms. The corresponding Gaussian smoothing was performed using $\sigma = 1$. FISTA and iPiano were implemented as

outlined in [Goldstein et al., 2014, Alg. 3] and [Ochs et al., 2014, Alg. 4], respectively, whereby second-order Tikhonov (TK2) regularization was applied using cubic B-Spline filtering for the proximal step. The TK2-regularization parameter $\lambda = 0.5$ was set based on the relationship $\sigma^2 = \sqrt{2\lambda}$ to the Gaussian standard deviation σ established experimentally in [Unser et al., 1993]. The step size parameter τ_0 was arbitrarily initialized as in [Goldstein et al., 2014] by ensuring a value higher than $2/L(\nabla f)$ from where backtracking line search took care of finding the adequate step size for the FBS variants. For iPiano the inertial weight β was set to 0.95 in conjunction with $c = 1.05$ and $\eta = 1.2$ to adaptively tune the step size parameter during runtime. The identity deformation was provided as the initial value for all solvers.

Figure 5.1 shows that both FISTA and iPiano outperform the demons and inertial demons algorithm in terms of computational speed and numerical accuracy converging about four times faster than their demons counterparts. The oscillations of FISTA at the beginning can be attributed to the non-monotone backtracking line search to ensure that iterates with higher objective values, but possibly closer to the minimizer, are not discarded.

5.3.2. 3D Anatomical MRI

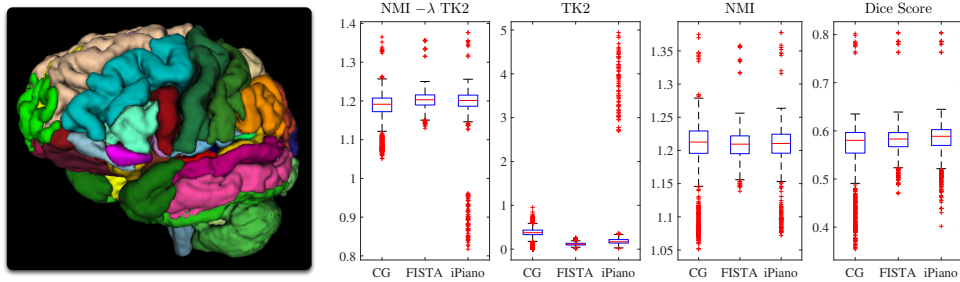


Figure 5.2.: 3D-Comparison of FISTA and iPiano against the conjugate gradient (CG) solver based on overall 1180 registrations. The Dice scores represent the mean of all propagated labels for each registration. All FBS results, apart from NMI, are statistically significantly better ($p < 10^{-6}$) than the CG ones based on the Wilcoxon signed rank test.

In the 3D experiment, we assessed the proposed FBS framework by registering 35 T1-weighted brain MRIs as provided by Neuromorphometrics for the MICCAI 2012

Grand Challenge on label fusion and propagated their respective parcellations holding 143 labels. Each brain was registered to the remaining 34 brains in either direction summing up to overall 1190 registrations for each method. In order to evaluate the performance of the actual numerical solvers in this task, we embedded both FISTA and iPiano as additional optimizers into the NIFTYREG registration framework [Modat et al., 2010] which is based on a cubic B-Spline parameterization. As similarity measure, we chose normalized mutual information. Including a TK2-regularization term in the objective function, the conjugate gradient (CG) solver could be applied to the overall cost function (5.1). This could then directly be compared against the performance of both FISTA and iPiano using cubic B-Spline filtering for the proximal operator. For the regularization parameter λ we chose 0.05. The step size for both FBS solvers was fixed and set experimentally to avoid its possible re-computation during the backtracking at each iteration and β was set to 0.95 for iPiano. The registration was performed within a multi-resolution framework with 2 levels. The FBS solvers ran for the total amount of 500 iterations whereas the CG solver was terminated earlier in case the stopping condition was met as implemented in NIFTYREG.

Figure 5.2 summarizes the corresponding registration results. Both FBS solvers reached statistically significantly better figures for overall similarity ($\text{NMI} - \lambda\text{TK2}$), smoothness (TK2), and Dice scores. However, 68 values were detected as outliers with higher values in the TK2 term for iPiano. Further analysis needs to be done but, presumably, this can be attributed to a possibly too high inertial weight β used in the experiment.

5.4. Discussion

In this chapter, we presented and demonstrated the capability of Forward-Backward Splitting schemes to be efficiently used in the challenging problem of deformable image registration. The advantage of FBS lies in their general framework to solve for arbitrary differentiable similarity measures in conjunction with any kind of convex regularizer associated with an easy-to-compute proximal operator. In addition,

the use of FBS comes with a proven algorithmic convergence to a critical value of the original problem (5.1). We provide explicit forward and proximal steps for several similarity and regularisation terms of benefit for (medical) image registration thereby allowing for efficient solutions via FBS. Additionally, and for the first time, we show the possibility to incorporate Tikhonov regularization in image registration by the mere application of B-Spline filtering in the proximal step which can be efficiently implemented by recursive IIR filters. We showcase two important instances of FBS solvers, FISTA and iPiano, and obtain statistically significantly better results in our numerical experiments than those obtained by the conjugate gradient method in NIFTYREG. Overall, we recommend the use of FBS methods to efficiently solve deformable registration problems due to their favorable theoretical properties, simplicity to use and general applicability to a wide range of similarity measures and regularizers. Future work could include the investigation of obtained deformation fields using FBS and a comparison with Quasi-Newton optimization methods. Moreover, FBS could constitute a more efficient means to compute the transformation parameters within the volumetric reconstruction frameworks. In particular, inter-slice regularization could be achieved by using Tikhonov regularization to impose a smoothness prior in between temporal neighboring slices for increased robustness in the slice-to-volume registration estimates.

Algorithm 5.3: FISTA with “restart” method, non-monotone backtracking line search and residual computation [Goldstein et al., 2014]. Residual computation is provided for completeness and was not incorporated to generate the results presented in here. Provided a differentiable, convex function f and convex (possibly non-smooth) function g , FISTA is guaranteed to converge.

Input : Starting point \mathbf{u}_0 , optional initial step size estimate τ_0
Output : Critical point estimate $\hat{\mathbf{u}}$ of $\arg \min_{\mathbf{u}} (f(\mathbf{u}) + g(\mathbf{u}))$

```

1  $\mathbf{v}_{-1} := \mathbf{u}_0, \mathbf{v}_0 := \mathbf{u}_0, \alpha_0 := 1, k := 0$ 
2 if initial step size estimate  $\tau_0$  is not provided then
3    $\tau_0 := 5 \frac{2}{L_{\text{est}}}$  with  $L_{\text{est}} := \frac{\|\nabla f(\check{\mathbf{u}}_1) - \nabla f(\check{\mathbf{u}}_2)\|}{\|\check{\mathbf{u}}_1 - \check{\mathbf{u}}_2\|} \leq L(\nabla f)$  for random  $\check{\mathbf{u}}_1, \check{\mathbf{u}}_2$ 
4 end
5 while residual  $r_k > \varepsilon_{\text{tol}}$  and  $k < k_{\text{max}}$  do
6    $\tilde{\mathbf{v}}_{k+1} := \mathbf{u}_k - \tau_k \nabla f(\mathbf{u}_k)$  (forward step)
7    $\mathbf{v}_{k+1} := \text{prox}_g(\tilde{\mathbf{v}}_{k+1}, \tau_k)$  (proximal step)
8    $\alpha_{k+1} := (1 + \sqrt{1 + 4\alpha_k^2})/2$  (acceleration parameter)
9    $\mathbf{u}_{k+1} := \mathbf{v}_{k+1} + \frac{\alpha_k - 1}{\alpha_{k+1}}(\mathbf{v}_{k+1} - \mathbf{v}_k)$  (prediction step)
10  while non-monotone backtracking line search condition
      
$$f(\mathbf{u}_{k+1}) \leq \hat{f}^k + \langle \nabla f(\mathbf{u}_k), \mathbf{u}_{k+1} - \mathbf{u}_k \rangle + \frac{1}{2\tau_k} \|\mathbf{u}_{k+1} - \mathbf{u}_k\|^2$$

      with  $\hat{f}^k := \max\{f(\mathbf{u}_{k-1}), f(\mathbf{u}_{k-2}), \dots, f(\mathbf{u}_{k-\min\{M, k\}})\}$ ,  $M \in \mathbb{N}$ , is not fulfilled  

      (to ensure  $\tau_k < 2/L(\nabla f)$ ) do
11     $\tau_k \leftarrow \tau_k/2$  (decrease step size)
12    recompute lines 6 to 9
13  end
14  begin
15    Compute residual  $r_{k+1} := \min\{r_{k+1}^r, r_{k+1}^n\}$  based on relative and normalized residuals
      
$$r_{k+1}^r := \frac{\|\hat{r}_{k+1}\|}{\max\{\|\nabla f(\mathbf{v}_{k+1})\|, \|\frac{\tilde{\mathbf{v}}_{k+1} - \mathbf{v}_{k+1}}{\tau_k}\|\} + \varepsilon^r}, \quad r_{k+1}^n := \frac{\|\hat{r}_{k+1}\|}{\hat{r}_1 + \varepsilon^n} \quad (5.9)$$

      with
      
$$\hat{r}_{k+1} := \nabla f(\mathbf{v}_{k+1}) + \frac{\tilde{\mathbf{v}}_{k+1} - \mathbf{v}_{k+1}}{\tau_k} \quad (5.10)$$

16  end
17  if  $\langle \mathbf{u}_{k+1} - \mathbf{v}_{k+1}, \mathbf{v}_{k+1} - \mathbf{v}_k \rangle \geq 0$  then
18     $\alpha_{k+1} := 1$  (“restart” method)
19  end
20   $k \leftarrow k + 1$ 
21 end
```

Algorithm 5.4: Non-monotone iPiano algorithm with backtracking [Ochs et al., 2014] and residual computation [Goldstein et al., 2014]. Residual computation is provided for completeness and was not incorporated to generate the results presented in here. Provided a differentiable (possibly non-convex) function f and convex (possibly non-smooth) function g , iPiano is guaranteed to converge.

Input : Starting point \mathbf{u}_0 , inertial weight $\beta \in [0, 1)$, step size tuning parameters $\eta > 1$ and $c > 1$ and optional initial Lipschitz-constant estimate $L_0 > 0$

Output : Critical point estimate $\hat{\mathbf{u}}$ of $\arg \min_{\mathbf{u}} (f(\mathbf{u}) + g(\mathbf{u}))$

```

1  $\mathbf{u}_{-1} := \mathbf{u}_0, k := 0$ 
2 if initial Lipschitz-constant estimate  $L_0$  is not provided then
3    $L_0 := \frac{L_{\text{est}}}{5}$  with  $L_{\text{est}} := \frac{\|\nabla f(\check{\mathbf{u}}_1) - \nabla f(\check{\mathbf{u}}_2)\|}{\|\check{\mathbf{u}}_1 - \check{\mathbf{u}}_2\|} \leq L(\nabla f)$  for random  $\check{\mathbf{u}}_1, \check{\mathbf{u}}_2$ 
4 end
5 while residual  $r_k > \varepsilon_{\text{tol}}$  and  $k < k_{\text{max}}$  do
6    $\tau_k := \frac{1.99(1-\beta)}{L_k}$ 
7    $\tilde{\mathbf{u}}_{k+1} := \mathbf{u}_k - \tau_k \nabla f(\mathbf{u}_k) + \beta(\mathbf{u}_k - \mathbf{u}_{k-1})$  (forward step with inertia)
8    $\mathbf{u}_{k+1} := \text{prox}_g(\tilde{\mathbf{u}}_{k+1}, \tau_k)$  (proximal step)
9   while monotone backtracking line search condition
      
$$f(\mathbf{u}_{k+1}) \leq f(\mathbf{u}_k) + \langle \nabla f(\mathbf{u}_k), \mathbf{u}_{k+1} - \mathbf{u}_k \rangle + \frac{L_k}{2} \|\mathbf{u}_{k+1} - \mathbf{u}_k\|^2$$

      is not fulfilled do
10     $L_k \leftarrow \eta L_k$  (increase Lipschitz-constant estimate)
11    recompute lines 6 to 8
12  end
13  Compute residual  $r_{k+1} = \min\{r_{k+1}^r, r_{k+1}^n\}$  via (5.9) and (5.10).
14   $L_{k+1} := L_k/c$  (decrease Lipschitz-constant estimate)
15   $k \leftarrow k + 1$ 
16 end
```

CHAPTER 6.

Conclusions and Future Work

Table of Contents

6.1. Conclusions	201
6.2. Discussion and Future Work	203

6.1. Conclusions

This thesis presents robust volumetric reconstruction frameworks from motion-corrupted, potentially sparse, 2D slices using inverse problem formulations applicable to a broad field of clinical applications in which associated motion patterns are inherently different, but the use of thick-slice MR data is current clinical practice.

In **Chapter 2**, an alternative to currently existing reconstruction frameworks for fetal brain MRI reconstruction was presented. A novel, complete outlier-robust super-resolution reconstruction framework was described that relies on a single hyperparameter only and retains a linear least-squares formulation, which can be solved efficiently. Despite its simplicity, it was shown to be an effective outlier-rejection approach to create high-fidelity high-resolution reconstructions with clear tissue boundary definitions even for very challenging motion- and intensity-artefact-corrupted low-resolution input data. A fast and automatic template-space alignment approach for automatic high-resolution visualization in the standard anatomical planes suitable for pathological brains was presented. Additionally, a robust motion-correction method based on Gaussian process regularization was presented that encourages the

consistency of spatial transformations with the overall aim of achieving more robust slice motion-correction estimates. The presented volumetric reconstruction framework was tested as part of an automatic localization, segmentation and volumetric reconstruction pipeline for fetal brain MRI, which generated high-resolution brain visualizations in the standard anatomical planes visually indistinguishable from those obtained by manual, labour-intensive brain segmentations. This framework unlocks the potential use of automatic fetal brain reconstruction studies in clinical practice.

In **Chapter 3**, super-resolution reconstruction frameworks were introduced to generate isotropic, high-resolution 3D visualizations of upper gastrointestinal anatomy from MRCP data. A reference-guided multimodal, deformable registration approach was presented for volumetric reconstruction in the context of sparse T2W MRI data available in current clinical MRCP protocols. By presenting a PSF-aware registration approach for motion correction, a consistent SRR framework was introduced that takes into account the physical image formation process for both registration and reconstruction steps. In a follow-up pilot study on healthy volunteers, the potential clinical utility of Super-Resolution was investigated for instances when current clinical MRCP protocols are extended by additional multiplanar SST2W stacks. Experiments showed the suitability and effectiveness of the proposed robust volumetric reconstruction framework initially presented for fetal brain MRI. In a controlled environment using quasi-static brain control data, the optimal number and orientation of SST2W series for MRCP SRR generation was empirically estimated. A comprehensive SRR quality assessment was conducted including four expert-readers with subspecialty interest in hepato-pancreatico-biliary imaging who assessed a range of anatomical regions particularly critical for the management of cancer patients. The results underline the potential of using SRR alongside traditional MRCP data for improved clinical diagnosis.

In **Chapter 4**, a volumetric reconstruction framework was presented that recovers geometrically consistent, volumetric images from printed films in MR neuroimaging. The flexible framework was based on semi-automatic slice extraction followed by automated slice-to-volume registration with inter-slice transformation regularization

and slice intensity correction whereby the missing meta-data information is contributed by a longitudinal scan of the same subject. A subsequent, isotropic in-plane total variation deconvolution technique was applied for visual enhancement of the obtained volumes. The validation of the proposed algorithm was performed on a uniquely long-term MRI dataset where a ground-truth was available. The results illustrated the robustness against numerous detrimental effects present in archaic films and the capability to accurately reconstruct 3D volumes. For follow-up research [Chung et al., 2019], the presented framework was used to reconstruct more than 200 image acquisitions captured on historical films to support a landmark longitudinal study¹ of 132 multiple sclerosis (MS) patients who had had been tracked since the 1980s. These reconstructions may reveal the full potential of modern image analysis tools and help understand disease progression in MS.

In **Chapter 5**, Forward-Backward Splitting (FBS) methods were presented to show their suitability in the context of deformable image registration and their potential advantage to efficiently achieve more robust slice-to-volume registration estimates for the proposed volumetric reconstruction frameworks. It was illustrated that for a range of cost functions vital for image registration, FBS variants can guarantee the convergence to a critical point by performing two simple, iterative steps. In particular, this framework guarantees that registration problems with any differentiable (not necessarily convex) similarity measure and convex (not necessarily smooth) regularization with a tractable proximal operator can be solved efficiently. The comparisons against state-of-the-art solvers performed on both synthetic and clinical data showed the advantage of FBS in terms of both convergence and registration accuracy.

6.2. Discussion and Future Work

The results presented in this thesis are encouraging, but the extent to which the proposed reconstruction methods will provide actual clinical utility and benefit in

¹<https://www.ucl.ac.uk/healthcare-engineering/news/2018/feb/image-reconstruction-tool-allows-researchers-unlock-historic-ms-data>

practice remains to be seen.

The developed fully-automated reconstruction framework for fetal brain MRI could be particularly useful in the context of spina bifida, in which MRI plays a role in characterizing the spinal lesion as well as resulting brain changes. The high-resolution reconstructions obtained by this framework may aid more accurate brain tissue measurement, which would help identify pathological changes associated with conditions such as spina bifida and other prognostically important brain changes. While the associated work in this thesis focused on methodological development and validation, a follow-up clinical study reconstructing more than 300 MRI examinations with the presented algorithm is currently under preparation. Moreover, the deployment of this algorithm for clinical use could be facilitated by its integration into GIFT-Cloud [Doel et al., 2017], a data sharing and collaboration platform between hospitals and research institutions. Such a large dataset will provide invaluable insight into the clinical applicability of the current algorithm. In particular, it may lead to revisiting some of the investigated robust motion-correction approaches presented in this thesis that had proven to be less effective in their current form (such as the GPR approach in Section 2.2.3) or to alternative initialization methods such as proposed in [Hou et al., 2018].

The same reconstruction algorithm also produced promising results in the reconstruction of the upper gastrointestinal anatomy of healthy subjects when an extended MRCP acquisition protocol is used that is not clinical standard. Future work aims at applying the proposed extended MRCP protocol to a representative patient cohort to assess the clinical utility of MRCP SRR. However, as data acquired during breath-holds of patients is expected to have more challenging motion artefacts, an algorithmic improvement may well be necessary. In particular, the currently used rigid motion correction model is inherently limited in capturing the complex deformations encountered in upper abdominal imaging. A poly-rigid (or poly-affine by simple extension) motion correction framework can be achieved by splitting a slice into smaller patches for a subsequent rigid patch-to-volume registration for each individual patch [Alansary et al., 2017]. However, it is non-trivial to define phys-

iologically plausible patches and the associated, reduced support of a single patch can complicate an accurate image registration, potentially leading to an increased rate of misregistrations. A promising approach could involve defining a robust joint-regularized motion-correction model, similar to the one presented in Section 4.2.2, which takes advantage of an additional 3D reference in (extended) MRCP protocols, such as the BFFE sequence. In general, a reference-guided approach appears particularly promising if a suitable multimodal similarity measure is available. Recently proposed deep learning methods [Cheng et al., 2018, Simonovsky et al., 2016] could prove useful for this step or, as shown in other applications, could be used to aid the motion correction as a whole [Hu et al., 2018]. Alternatively, recent deep learning approaches show promising results to synthesize images using image-to-image translation [Zhu et al., 2017, Isola et al., 2017, Wolterink et al., 2017]. In particular, pseudo-monomodal registration could be performed in case a suitable, same-contrast reference volume is synthesized [Tanner et al., 2018, Hiasa et al., 2018]. Furthermore, learning-based reconstruction methods could be another avenue worth exploring. However, key challenges would include the need for a sufficiently large training dataset of MRCP studies in addition to a potential poor ability to generalize in cases of complex, and potentially, substantial motion in abdominal anatomy (Figure 1.2).

Forward-Backward Splitting approaches represent an attractive numerical scheme when a tractable proximal operator is available. If this leads to an overall smooth cost function, such as given by a Tikhonov regularizer (Section 5.2.3), their advantage against Quasi-Newton methods is yet to be demonstrated. However, a potentially much more clinically interesting application for FBS could be found in deformable image registration constrained by physiologically relevant tissue properties such as elasticity or incompressibility [Mansi et al., 2011]. Imposing such constraints leads to an overall non-smooth cost function that in practice is difficult to solve using traditional solvers. In the framework of FBS, however, the associated incompressibility term results in a tractable proximal operator. Thus, using FBS could prove particularly helpful in the context of *incompressible* image registration to achieve physiologically plausible transformations [Fidon et al., 2019].

Appendices

APPENDIX A.

Super-Resolution Reconstruction of Fetal Brain MRI

A.1. Evaluation of Volumetric Self-Consistency

In addition to the performed slice similarity comparisons $\text{Sim}(\mathbf{y}_k^i, \mathbf{A}_k^i \mathbf{x}^i)$ after the final SVR-SRR iteration ($i = 3$) using SSIM and PSNR in Figure 2.16, we also performed comparisons using NCC, normalised mutual information (NMI), root mean squared error (RMSE) and mean absolute error (MAE). Figure A.1 summarizes the comparisons which corroborate the findings using SSIM and PSNR that SRR (S)/(M) appear of similar volumetric self-consistency.

A.2. Clinical Evaluation

Figures A.2 and A.3 represent an extension to Figure 2.17 and provide a more detailed comparison of the individual scores regarding anatomical clarity and SRR quality.

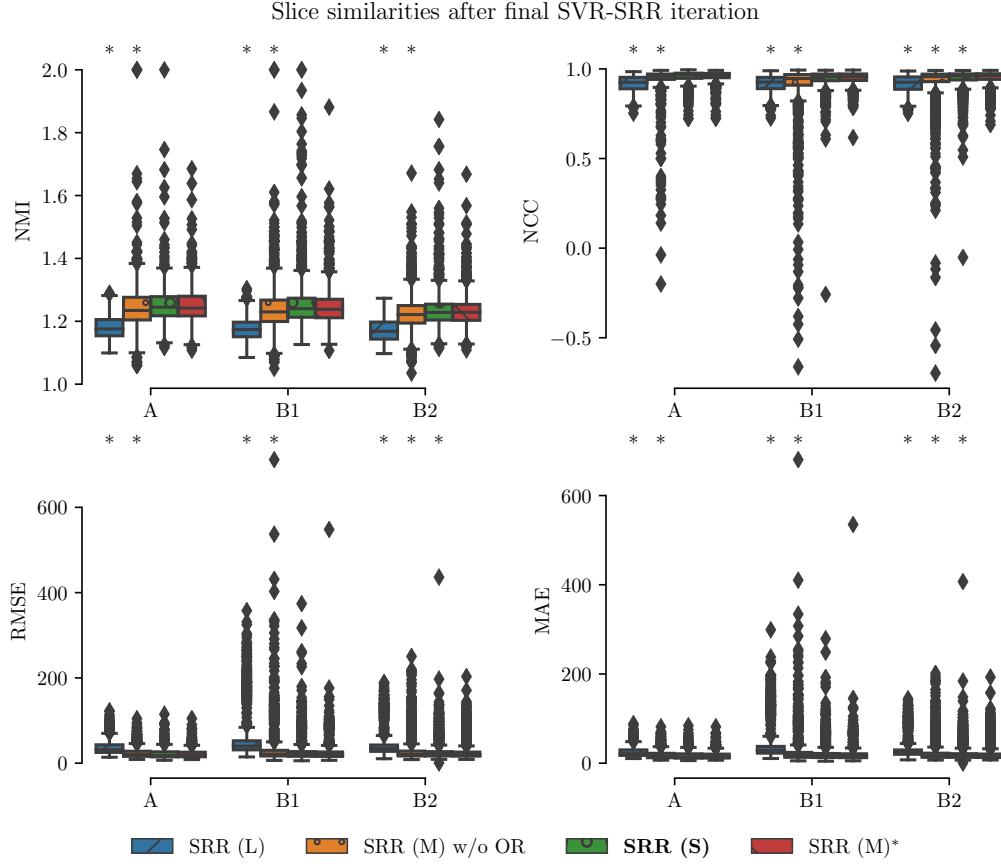


Figure A.1.: Slice similarities after the final SVR-SRR iteration. Quantitative comparison of different reconstruction methods based on $\text{Sim}(\mathbf{y}_k^i, \mathbf{A}_k^i \mathbf{x}^i)$ after the final SVR-SRR iteration ($i = 3$). A * denotes a significant difference compared to SRR (M) within each group based on Kruskal-Wallis with post-hoc Dunn tests ($p < 0.05$). Thus, SRR (S) and SRR (M) appear of similar volumetric self-consistency as quantified by the similarities between motion-corrected and respectively projected HR volume slices.

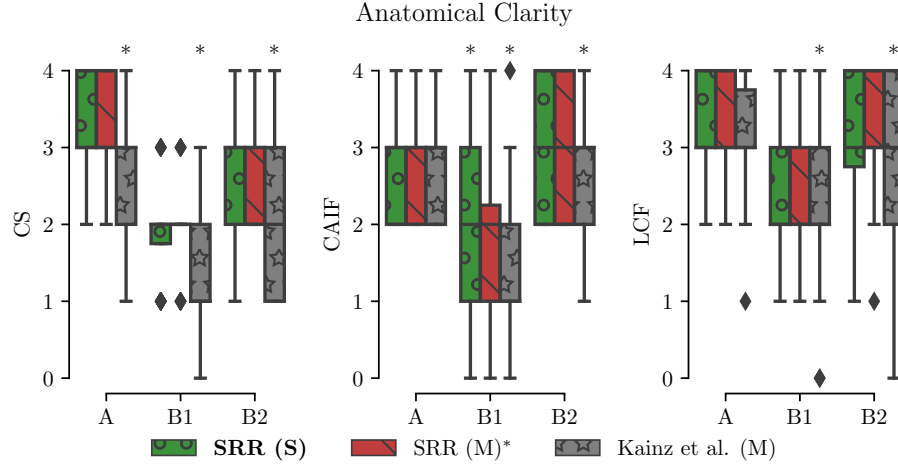


Figure A.2.: Summary of clinical evaluation for anatomical clarity scores. Two radiologists performed a qualitative assessment of the obtained HR reconstructions regarding anatomical clarity involving 39 cases. Scores indicate how well cerebellar structure (CS), cerebral aqueduct and interhemispheric fissure (CAIF) and longitudinal cerebral fissure (LCF) are visualized in each image with ratings 0 (structure not seen), 1 (poor depiction), 2 (suboptimal visualization; image not adequate for diagnostic purposes), 3 (clear visualization of structure but reduced tissue contrast; image-based diagnosis feasible), and 4 (excellent depiction; optimal for diagnostic purposes). A * denotes a significant difference compared to SRR (M) based on a Wilcoxon signed-rank test ($p < 0.05$).

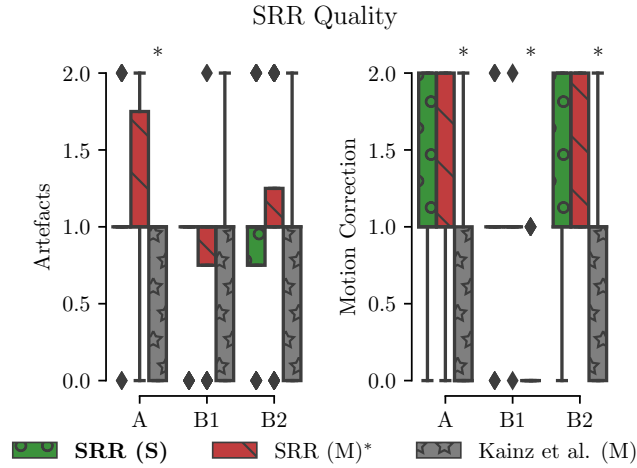


Figure A.3.: Summary of clinical evaluation for SRR quality scores. Two radiologists performed a qualitative assessment of the obtained HR reconstructions regarding SRR quality involving 39 cases. The SRR quality was described by the visible artefacts and blur scores with ratings 0 (lots of artefacts/blur) to 2 (no artefact/blur). A * denotes a significant difference compared to SRR (M) based on a Wilcoxon signed-rank test ($p < 0.05$).

APPENDIX B.

Super-Resolution Reconstruction of Upper Abdominal MRI

B.1. Optimisation Control Studies for Brain MRI SRR

The qualitative comparison in Figure B.1 extends the one in Figure 3.8 by providing additional axial and coronal views for both static and motion-corrected brain SRR outcomes.

Figures B.2 and B.3 show SSIM, PSNR and NMI in addition to NCC as provided in Figure 3.7.

Table B.1 provides an extension to Table 3.3 for more source data configurations, additional similarity measures and the axial SST2W stack (SST2W_{ax}) as another possible choice as reference image for the reference-guided SRR approach. Using the short-hand "RG-Reference-SimilarityMeasureForRegistration", the settings for the reference-guided SRR approach are shown, where, e.g., RG-BFFE-NMI refers to the use of BFFE as reference volume for guidance and NMI as similarity measure for registration. Using NMI, as shown for the two references of HT2W and the axial SST2W stack in Table 3.3, can be computationally unstable and eventually fail as voxel numbers can be insufficient for the slice-to-volume metric evaluations.

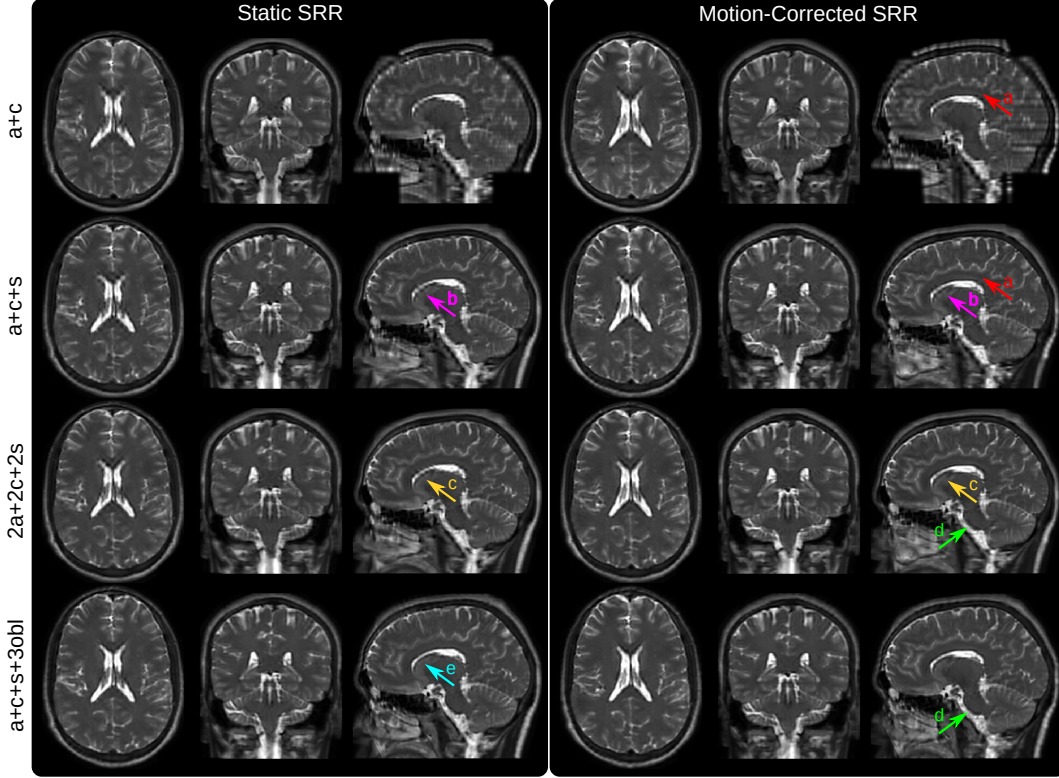


Figure B.1.: Qualitative comparison of the static and reference-guided SRR outcome of one subject for various input data scenarios (extension to Figure 3.8). It illustrates the impact of the number of input stacks and how multiple orientations can improve PVE recovery. In particular, SRR ($a+c+s+3obl$) shows visually higher anatomical accuracy than SRR ($2a+2c+2s$) despite the same number of six input stacks used for the SRR. The red arrows (a) underline that the SRR based on only two stacks ($a+c$) as currently available for clinical MRCP study protocols produces a very poor SRR quality which is especially noticeable in the sagittal view. The magenta arrows (b) illustrate that for three input stacks ($a+c+s$) the corpus callosum can only be reconstructed with limited geometrical integrity. Motion-correction helps to recover it more clearly by adding three additional stacks ($2a+2c+2s$) as indicated by arrows (c). The green arrows (d) show the improved visual clarity at the medulla due to better PVE correction by using oblique data. Additional oversampling for high input stack numbers leads to higher PSNR. This may also result in clear tissue boundaries even in case of insufficient motion correction for the static SRR as indicated by the cyan arrow (e).

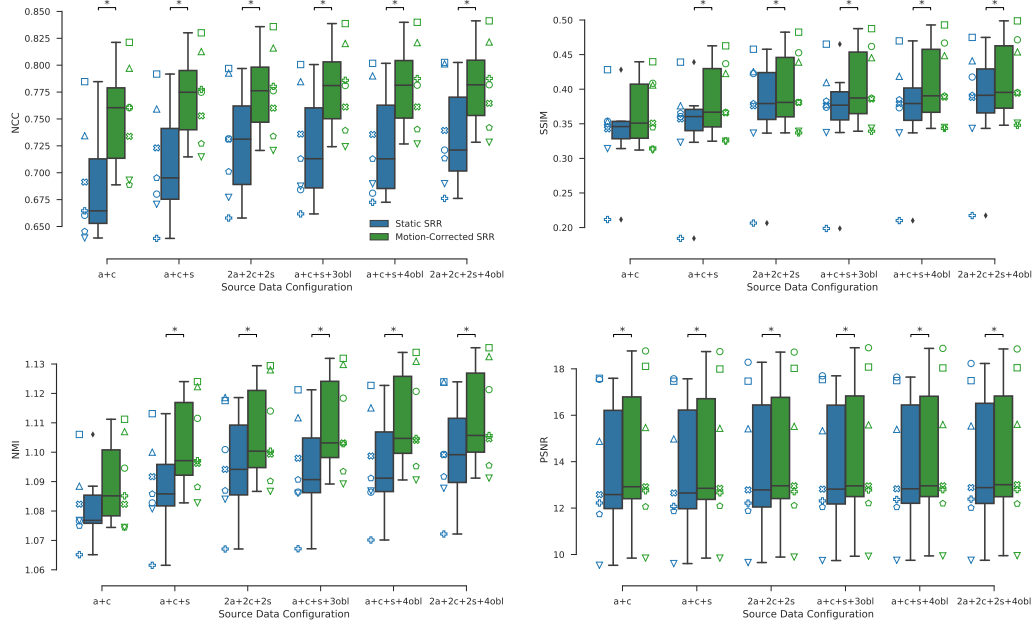


Figure B.2.: Ground-truth (HR T2W) similarities for static and reference-guided SRR outcomes for the quasi-static brain experiment involving seven subjects. The more input stacks are used the higher the similarity scores. Moreover, motion correction markedly improves the ground-truth similarities which was performed by rigidly registering each individual slice to the HR T2W volume using NCC as the similarity measure. Stars indicate statistical differences between the groups using a pairwise Wilcoxon signed-rank test ($p < 0.05$).

B.2. Optimisation Studies for Upper Abdominal MRI SRR

Figure B.4 shows SSIM, PSNR and NMI in addition to NCC as provided in Figure 3.9. Table B.3 provides a numerical summary of investigated motion-correction strategies including the reference-guided approaches without the in-plane deformation step (RigidOnly).

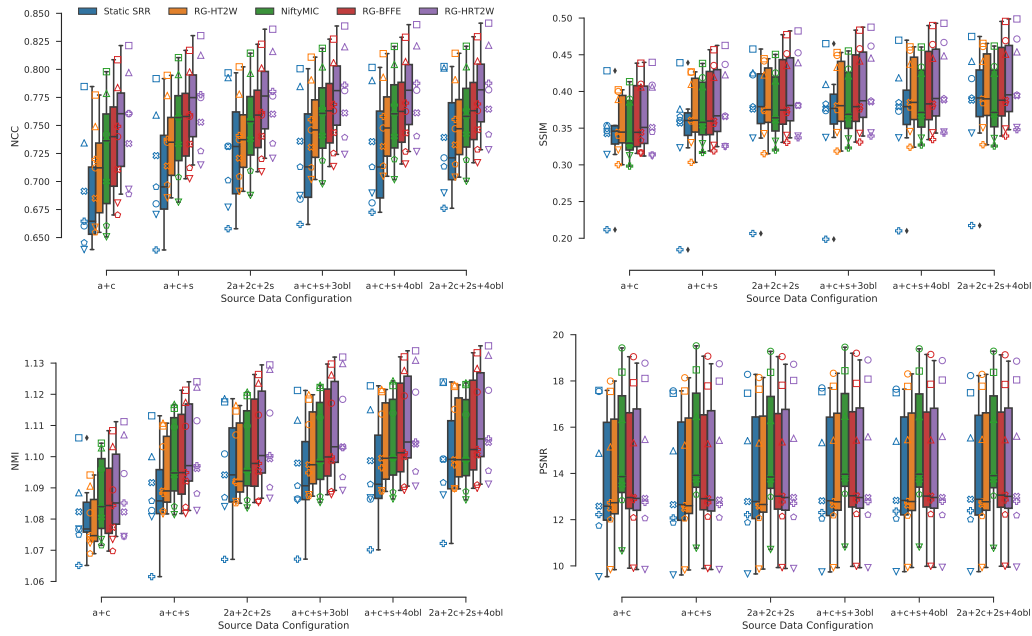


Figure B.3.: Ground-truth (HR T2W) similarities for the quasi-static brain experiment for all registration/motion correction strategies as an extension to Figure B.2. Reference-guided approaches used NCC as the similarity measure for registration.

Table B.1.: Ground-truth (HR T2W) similarities of obtained quasi-static control brain SRRs for an increasing number of input stacks for different motion correction (MC) strategies summarized for all seven subjects. The rows are sorted in a descending order according to the SRR outcome for 'a+c+s+3obl'. Using the short-hand "RG-Reference-SimilarityMeasureForRegistration", the settings for the reference-guided SRR approach are shown, where, e.g., RG-BFFE-NMI refers to the use of BFFE as reference volume for guidance and NMI as similarity measure for registration.

(a) NCC						
MC Strategy	a+c	a+c+s	2a+2c+2s	a+c+s+3obl	a+c+s+4obl	2a+2a+2c+4obl
RG-HRT2W-NCC	0.751 ± 0.046	0.770 ± 0.039	0.775 ± 0.038	0.779 ± 0.038	0.780 ± 0.038	0.781 ± 0.038
RG-HRT2W-MI	0.722 ± 0.050	0.760 ± 0.040	0.772 ± 0.038	0.777 ± 0.038	0.780 ± 0.037	0.784 ± 0.037
RG-HRT2W-NMI	0.738 ± 0.051	0.760 ± 0.043	0.764 ± 0.042	0.771 ± 0.041	0.771 ± 0.041	0.772 ± 0.041
RG-BFFE-NCC	0.735 ± 0.047	0.754 ± 0.039	0.759 ± 0.038	0.764 ± 0.038	0.765 ± 0.038	0.766 ± 0.038
RG-BFFE-NMI	0.726 ± 0.052	0.748 ± 0.043	0.751 ± 0.042	0.758 ± 0.041	0.760 ± 0.041	0.759 ± 0.041
NiftyMIC	0.724 ± 0.052	0.748 ± 0.043	0.751 ± 0.041	0.758 ± 0.040	0.759 ± 0.040	0.758 ± 0.040
RG-HT2W-NCC	0.708 ± 0.042	0.734 ± 0.037	0.739 ± 0.037	0.750 ± 0.037	0.752 ± 0.037	0.751 ± 0.037
RG-SST2Wax-NCC	0.706 ± 0.049	0.717 ± 0.042	0.724 ± 0.041	0.735 ± 0.041	0.739 ± 0.040	0.739 ± 0.043
Static SRR	0.689 ± 0.049	0.708 ± 0.049	0.727 ± 0.050	0.724 ± 0.049	0.726 ± 0.048	0.735 ± 0.046
RG-BFFE-MI	0.353 ± 0.176	0.389 ± 0.196	0.427 ± 0.190	0.484 ± 0.183	0.506 ± 0.178	0.525 ± 0.168
RG-SST2Wax-MI	0.295 ± 0.052	0.336 ± 0.059	0.384 ± 0.053	0.414 ± 0.052	0.434 ± 0.048	0.463 ± 0.046
RG-HT2W-MI	0.242 ± 0.074	0.258 ± 0.079	0.310 ± 0.092	0.288 ± 0.094	0.290 ± 0.092	0.336 ± 0.101
(b) SSIM						
MC Strategy	a+c	a+c+s	2a+2c+2s	a+c+s+3obl	a+c+s+4obl	2a+2a+2c+4obl
RG-HRT2W-NMI	0.378 ± 0.045	0.394 ± 0.050	0.408 ± 0.052	0.411 ± 0.054	0.414 ± 0.055	0.419 ± 0.055
RG-HRT2W-NCC	0.368 ± 0.047	0.386 ± 0.051	0.402 ± 0.053	0.407 ± 0.054	0.411 ± 0.055	0.416 ± 0.055
RG-BFFE-NMI	0.372 ± 0.046	0.386 ± 0.051	0.400 ± 0.054	0.404 ± 0.056	0.407 ± 0.057	0.412 ± 0.057
RG-BFFE-NCC	0.366 ± 0.047	0.382 ± 0.051	0.397 ± 0.053	0.402 ± 0.055	0.406 ± 0.056	0.411 ± 0.056
RG-MRCP-NCC	0.357 ± 0.038	0.374 ± 0.045	0.388 ± 0.046	0.393 ± 0.049	0.398 ± 0.050	0.403 ± 0.050
RG-HRT2W-MI	0.336 ± 0.055	0.365 ± 0.052	0.385 ± 0.052	0.391 ± 0.053	0.397 ± 0.053	0.405 ± 0.053
NiftyMIC	0.351 ± 0.041	0.374 ± 0.043	0.381 ± 0.046	0.385 ± 0.047	0.389 ± 0.047	0.389 ± 0.048
RG-SST2Wax-NCC	0.356 ± 0.043	0.358 ± 0.044	0.373 ± 0.047	0.382 ± 0.049	0.389 ± 0.050	0.388 ± 0.050
Static SRR	0.336 ± 0.060	0.344 ± 0.073	0.372 ± 0.077	0.364 ± 0.076	0.368 ± 0.075	0.382 ± 0.078
RG-BFFE-MI	0.099 ± 0.108	0.115 ± 0.130	0.129 ± 0.141	0.145 ± 0.151	0.154 ± 0.155	0.159 ± 0.157
RG-SST2Wax-MI	0.087 ± 0.022	0.086 ± 0.018	0.105 ± 0.018	0.101 ± 0.020	0.107 ± 0.021	0.121 ± 0.022
RG-MRCP-MI	0.059 ± 0.022	0.069 ± 0.022	0.084 ± 0.028	0.066 ± 0.026	0.065 ± 0.025	0.080 ± 0.028
(c) NMI						
MC Strategy	a+c	a+c+s	2a+2c+2s	a+c+s+3obl	a+c+s+4obl	2a+2a+2c+4obl
RG-HRT2W-NMI	1.092 ± 0.013	1.104 ± 0.015	1.108 ± 0.016	1.111 ± 0.016	1.112 ± 0.016	1.113 ± 0.017
RG-HRT2W-NCC	1.090 ± 0.014	1.103 ± 0.015	1.107 ± 0.016	1.110 ± 0.016	1.111 ± 0.016	1.112 ± 0.016
RG-BFFE-NMI	1.087 ± 0.013	1.100 ± 0.015	1.104 ± 0.016	1.107 ± 0.016	1.109 ± 0.016	1.109 ± 0.017
RG-BFFE-NCC	1.087 ± 0.014	1.100 ± 0.015	1.104 ± 0.016	1.107 ± 0.016	1.109 ± 0.016	1.109 ± 0.017
RG-HRT2W-MI	1.078 ± 0.015	1.096 ± 0.015	1.101 ± 0.016	1.104 ± 0.015	1.107 ± 0.016	1.108 ± 0.016
NiftyMIC	1.088 ± 0.013	1.099 ± 0.013	1.102 ± 0.014	1.104 ± 0.014	1.105 ± 0.014	1.105 ± 0.014
RG-MRCP-NCC	1.079 ± 0.009	1.095 ± 0.012	1.099 ± 0.012	1.102 ± 0.013	1.104 ± 0.013	1.105 ± 0.014
RG-SST2Wax-NCC	1.087 ± 0.012	1.091 ± 0.012	1.095 ± 0.013	1.099 ± 0.013	1.101 ± 0.014	1.101 ± 0.015
Static SRR	1.081 ± 0.012	1.088 ± 0.015	1.096 ± 0.017	1.094 ± 0.017	1.096 ± 0.017	1.100 ± 0.017
RG-BFFE-MI	1.022 ± 0.023	1.027 ± 0.031	1.031 ± 0.034	1.039 ± 0.037	1.043 ± 0.038	1.044 ± 0.038
RG-SST2Wax-MI	1.013 ± 0.004	1.015 ± 0.005	1.018 ± 0.004	1.021 ± 0.005	1.023 ± 0.004	1.026 ± 0.004
RG-MRCP-MI	1.009 ± 0.006	1.010 ± 0.006	1.013 ± 0.007	1.012 ± 0.006	1.012 ± 0.006	1.015 ± 0.007
(d) PSNR						
MC Strategy	a+c	a+c+s	2a+2c+2s	a+c+s+3obl	a+c+s+4obl	2a+2a+2c+4obl
NiftyMIC	15.001 ± 2.926	15.135 ± 2.902	15.030 ± 2.847	15.148 ± 2.863	15.123 ± 2.850	15.061 ± 2.832
RG-HRT2W-MI	14.162 ± 2.938	14.252 ± 2.990	14.398 ± 3.013	14.474 ± 3.031	14.498 ± 3.027	14.546 ± 3.043
RG-BFFE-MI	12.536 ± 2.219	13.328 ± 2.339	13.903 ± 2.337	14.412 ± 2.505	14.644 ± 2.541	14.904 ± 2.513
RG-BFFE-NCC	14.302 ± 3.043	14.257 ± 3.034	14.298 ± 3.020	14.378 ± 3.042	14.364 ± 3.025	14.365 ± 3.019
RG-HRT2W-NCC	14.275 ± 3.048	14.230 ± 3.027	14.276 ± 3.011	14.349 ± 3.042	14.344 ± 3.027	14.350 ± 3.020
RG-BFFE-NMI	14.243 ± 3.062	14.219 ± 3.048	14.244 ± 3.026	14.330 ± 3.051	14.313 ± 3.034	14.305 ± 3.025
RG-HRT2W-NMI	14.148 ± 3.049	14.129 ± 3.025	14.164 ± 3.007	14.251 ± 3.029	14.236 ± 3.011	14.232 ± 3.003
RG-MRCP-NCC	13.968 ± 2.803	13.984 ± 2.849	14.039 ± 2.850	14.151 ± 2.882	14.148 ± 2.877	14.150 ± 2.875
RG-SST2Wax-NCC	13.929 ± 2.983	13.906 ± 2.940	13.985 ± 2.936	14.130 ± 2.975	14.159 ± 2.967	13.426 ± 2.499
Static SRR	13.727 ± 2.827	13.746 ± 2.794	13.957 ± 2.938	13.926 ± 2.782	13.932 ± 2.759	14.041 ± 2.886
RG-SST2Wax-MI	11.822 ± 1.401	12.659 ± 1.686	13.437 ± 1.829	13.852 ± 1.976	14.201 ± 2.073	14.565 ± 2.158
RG-MRCP-MI	11.754 ± 1.369	12.658 ± 1.608	13.596 ± 1.824	13.529 ± 1.873	13.752 ± 1.915	14.239 ± 1.967

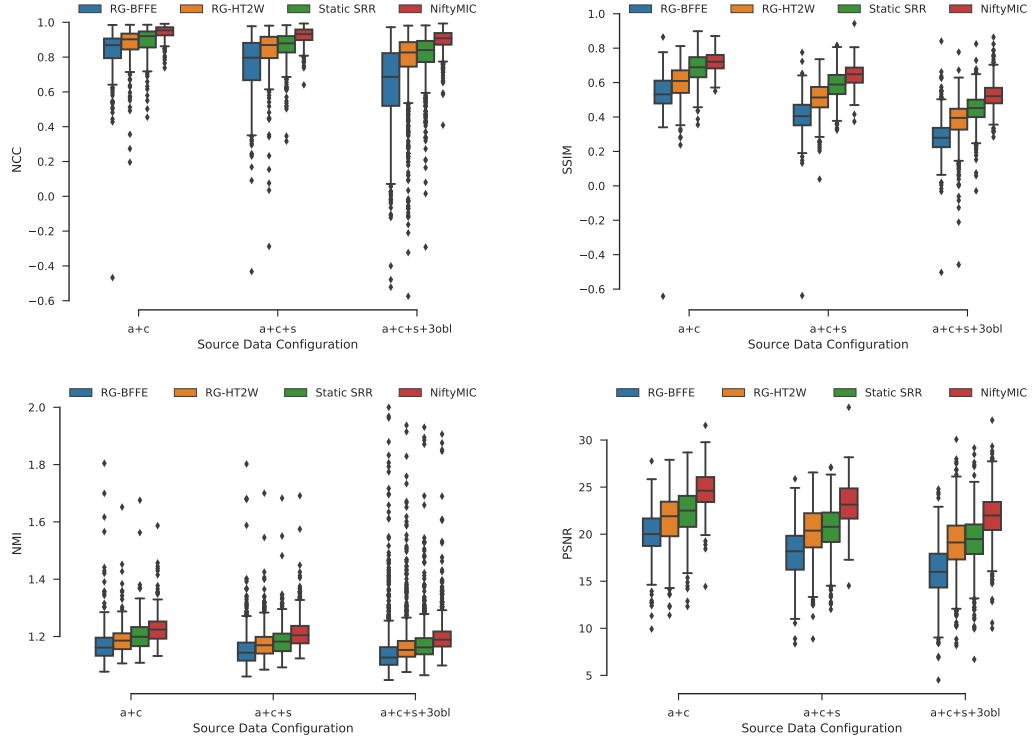


Figure B.4.: Self-consistency evaluation given by projected similarities for all slices of obtained abdominal SRRs for an increasing number of input stacks for different motion-correction strategies summarised for all eight subjects. All self-consistency outcomes between SRR approaches, except for RG-HT2W vs Static SRR for 'a+c+s', are significantly different within each source data configuration based on Kruskal-Wallis with post-hoc Dunn tests ($p < 0.05$).

Table B.3.: Projected slice similarity evaluation of obtained abdominal SRRs for an increasing number of input stacks for different motion correction strategies summarized for all eight subjects. The rows are sorted in a descending order according to the NCC/NMI-outcome for "a+c+s+3obl". NiftyMIC shows superior self-consistency across different number of input data scenarios.

(a) NCC and SSIM

MC Strategy	NCC			SSIM		
	a+c	a+c+s	a+c+s+3obl	a+c	a+c+s	a+c+s+3obl
NiftyMIC	0.942 ± 0.041	0.922 ± 0.050	0.898 ± 0.061	0.720 ± 0.056	0.645 ± 0.066	0.526 ± 0.072
Static SRR	0.892 ± 0.084	0.858 ± 0.093	0.812 ± 0.129	0.680 ± 0.089	0.584 ± 0.086	0.447 ± 0.085
RG-HT2W-NCC (RigidOnly)	0.872 ± 0.120	0.835 ± 0.144	0.778 ± 0.193	0.602 ± 0.109	0.511 ± 0.098	0.390 ± 0.104
RG-HT2W-NCC	0.871 ± 0.110	0.835 ± 0.138	0.772 ± 0.201	0.599 ± 0.102	0.506 ± 0.095	0.381 ± 0.107
RG-BFFE-NCC (RigidOnly)	0.834 ± 0.114	0.751 ± 0.172	0.651 ± 0.226	0.541 ± 0.100	0.414 ± 0.094	0.285 ± 0.095
RG-BFFE-NCC	0.832 ± 0.136	0.751 ± 0.179	0.648 ± 0.229	0.537 ± 0.122	0.409 ± 0.108	0.282 ± 0.098
RG-HT2W-MI	0.748 ± 0.158	0.687 ± 0.193	0.577 ± 0.248	0.500 ± 0.151	0.385 ± 0.115	0.257 ± 0.103
RG-HT2W-MI (RigidOnly)	0.749 ± 0.151	0.678 ± 0.171	0.565 ± 0.218	0.498 ± 0.146	0.372 ± 0.103	0.247 ± 0.092
RG-BFFE-MI	0.713 ± 0.188	0.634 ± 0.208	0.509 ± 0.233	0.446 ± 0.136	0.333 ± 0.108	0.215 ± 0.092
RG-BFFE-MI (RigidOnly)	0.706 ± 0.190	0.627 ± 0.201	0.505 ± 0.216	0.441 ± 0.138	0.334 ± 0.102	0.217 ± 0.085
RG-HT2W-NMI	—	—	—	—	—	—
RG-HT2W-NMI (RigidOnly)	—	—	—	—	—	—
RG-BFFE-NMI	—	—	—	—	—	—
RG-BFFE-NMI (RigidOnly)	—	—	—	—	—	—

(b) NMI and PSNR

MC Strategy	NMI			PSNR		
	a+c	a+c+s	a+c+s+3obl	a+c	a+c+s	a+c+s+3obl
NiftyMIC	1.230 ± 0.058	1.215 ± 0.061	1.203 ± 0.076	24.631 ± 2.170	23.217 ± 2.251	21.903 ± 2.434
Static SRR	1.205 ± 0.062	1.188 ± 0.059	1.180 ± 0.083	22.229 ± 2.689	20.631 ± 2.566	19.356 ± 2.546
RG-HT2W-NCC	1.192 ± 0.060	1.179 ± 0.062	1.174 ± 0.092	21.493 ± 2.927	20.160 ± 2.716	18.975 ± 2.893
RG-HT2W-NCC (RigidOnly)	1.194 ± 0.067	1.180 ± 0.065	1.174 ± 0.089	21.617 ± 3.003	20.250 ± 2.807	19.116 ± 2.846
RG-HT2W-MI	1.175 ± 0.099	1.162 ± 0.100	1.162 ± 0.134	17.818 ± 3.646	16.453 ± 3.333	14.918 ± 3.360
RG-BFFE-MI	1.175 ± 0.103	1.168 ± 0.112	1.161 ± 0.121	17.031 ± 3.119	15.417 ± 3.086	13.680 ± 2.789
RG-HT2W-MI (RigidOnly)	1.181 ± 0.127	1.163 ± 0.115	1.159 ± 0.140	17.721 ± 3.613	16.169 ± 3.246	14.530 ± 3.180
RG-BFFE-MI (RigidOnly)	1.171 ± 0.094	1.168 ± 0.108	1.159 ± 0.119	16.940 ± 2.924	15.269 ± 2.560	13.496 ± 2.517
RG-BFFE-NCC	1.181 ± 0.087	1.162 ± 0.083	1.156 ± 0.111	20.036 ± 2.583	18.020 ± 2.694	16.107 ± 2.751
RG-BFFE-NCC (RigidOnly)	1.181 ± 0.087	1.163 ± 0.086	1.156 ± 0.111	20.129 ± 2.436	18.090 ± 2.555	16.216 ± 2.649
RG-HT2W-NMI	—	—	—	—	—	—
RG-HT2W-NMI (RigidOnly)	—	—	—	—	—	—
RG-BFFE-NMI	—	—	—	—	—	—
RG-BFFE-NMI (RigidOnly)	—	—	—	—	—	—

Bibliography

- [Aertsen et al., 2019] Aertsen, M., Verduyckt, J., De Keyzer, F., Vercauteren, T., Van Calenbergh, F., De Catte, L., Dymarkowski, S., Demaerel, P., and Deprest, J. (2019). Reliability of MR Imaging-Based Posterior Fossa and Brain Stem Measurements in Open Spinal Dysraphism in the Era of Fetal Surgery. *American Journal of Neuroradiology*, 40(1):191–198.
- [Aganj et al., 2012] Aganj, I., Lenglet, C., Yacoub, E., Sapiro, G., and Harel, N. (2012). A 3D Wavelet Fusion Approach for the Reconstruction of Isotropic-Resolution MR Images From Orthogonal Anisotropic-Resolution Scans. *Magnetic Resonance in Medicine*, 67(4):1167–1172.
- [Akbar et al., 2018] Akbar, A., Tran, Q. T., Nair, S. P., Parikh, S., Bilal, M., Ismail, M., Vanatta, J. M., Eason, J. D., and Satapathy, S. K. (2018). Role of MRCP in Diagnosing Biliary Anastomotic Strictures After Liver Transplantation. *Transplantation Direct*, 4(5):e347.
- [Alansary et al., 2017] Alansary, A., Rajchl, M., McDonagh, S. G., Murgasova, M., Damodaram, M., Lloyd, D. F. A., Davidson, A., Rutherford, M., Hajnal, J. V., Rueckert, D., and Kainz, B. (2017). PVR: Patch-to-Volume Reconstruction for Large Area Motion Correction of Fetal MRI. *IEEE Transactions on Medical Imaging*, 36(10):2031–2044.
- [Altun et al., 2016] Altun, E., Velloni, F., AlObaidy, M., Elias, J., and Semelka, R. C. (2016). Gallbladder and biliary system. In *Abdominal-Pelvic MRI*, pages 395–459. John Wiley & Sons, Ltd, Oxford, UK, fourth edition.

- [Anquez et al., 2009] Anquez, J., Angelini, E. D., and Bloch, I. (2009). Automatic segmentation of head structures on fetal MRI. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 109–112. IEEE.
- [Ashburner, 2007] Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, 38(1):95–113.
- [Ashburner and Friston, 2005] Ashburner, J. and Friston, K. J. (2005). Unified segmentation. *NeuroImage*, 26(3):839–851.
- [Aydelotte, 2016] Aydelotte, J. (2016). Magnetic Resonance Cholangiopancreatography Is No Longer Necessary. *Journal of the American College of Surgeons*, 222(3):329.
- [Aydelotte et al., 2015] Aydelotte, J. D., Ali, J., Huynh, P. T., Coopwood, T. B., Uecker, J. M., and Brown, C. V. (2015). Use of Magnetic Resonance Cholangiopancreatography in Clinical Practice: Not as Good as We Once Thought. *Journal of the American College of Surgeons*, 221(1):215–219.
- [Baker and Kanade, 2002] Baker, S. and Kanade, T. (2002). Limits on Super-Resolution and How to Break Them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183.
- [Barish et al., 1999] Barish, M. A., Yucel, E. K., and Ferrucci, J. T. (1999). Magnetic Resonance Cholangiopancreatography. *New England Journal of Medicine*, 341(4):258–264.
- [Basaran et al., 2008] Basaran, C., Agildere, A. M., Donmez, F. Y., Sevmis, S., Budakoglu, I., Karakayali, H., and Haberal, M. (2008). MR Cholangiopancreatography with T2-Weighted Prospective Acquisition Correction Turbo Spin-Echo Sequence of the Biliary Anatomy of Potential Living Liver Transplant Donors. *American Journal of Roentgenology*, 190(6):1527–1533.
- [Basty and Grau, 2018] Basty, N. and Grau, V. (2018). Super Resolution of Cardiac Cine MRI Sequences Using Deep Learning. In Stoyanov, D., Taylor, Z., Kainz, B.,

- Maicas, G., Beichel, R. R., Martel, A., Maier-Hein, L., Bhatia, K., Vercauteren, T., Oktay, O., Carneiro, G., Bradley, A. P., Nascimento, J., Min, H., Brown, M. S., Jacobs, C., Lassen-Schmidt, B., Mori, K., Petersen, J., San José Estépar, R., Schmidt-Richberg, A., and Veiga, C., editors, *Image Analysis for Moving Organ, Breast, and Thoracic Images. RAMBO 2018, BIA 2018, TIA 2018*, volume 11040, pages 23–31. Springer, Cham.
- [Basty et al., 2017] Bastý, N., McClymont, D., Teh, I., Schneider, J. E., and Grau, V. (2017). Reconstruction of 3D Cardiac MR Images from 2D Slices Using Directional Total Variation. In Cardoso, M. J., Arbel, T., Gao, F., Kainz, B., van Walsum, T., Shi, K., Bhatia, K. K., Peter, R., Vercauteren, T., Reyes, M., Dalca, A., Wiest, R., Niessen, W., and Emmer, B. J., editors, *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment. RAMBO 2017, CMMI 2017, SWITCH 2017*, volume 10555 of *Lecture Notes in Computer Science*, pages 127–135, Cham. Springer, Cham.
- [Beck and Teboulle, 2009a] Beck, A. and Teboulle, M. (2009a). A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202.
- [Beck and Teboulle, 2009b] Beck, A. and Teboulle, M. (2009b). Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transactions on Image Processing*, 18(11):2419–2434.
- [Becker and Fadili, 2012] Becker, S. and Fadili, M. J. (2012). A quasi-Newton proximal splitting method. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 25, pages 2618–2626. Curran Associates, Inc.
- [Benson and Bluth, 2008] Benson, C. B. and Bluth, E. I. (2008). *Ultrasonography in Obstetrics and Gynecology: A Practical Approach to Clinical Problems*. Thieme Medical Publishers, Inc., Stuttgart, second edition.
- [Bhatia et al., 2014] Bhatia, K. K., Price, A. N., Shi, W., Hajnal, J. V., and Rueck-

- ert, D. (2014). Super-resolution reconstruction of cardiac MRI using coupled dictionary learning. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 947–950. IEEE.
- [Boyd et al., 2011] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122.
- [Brex et al., 2002] Brex, P. A., Ciccarelli, O., O’Riordan, J. I., Sailer, M., Thompson, A. J., and Miller, D. H. (2002). A Longitudinal Study of Abnormalities on MRI and Disability from Multiple Sclerosis. *New England Journal of Medicine*, 346(3):158–164.
- [Brix et al., 2016] Brix, L., Ringgaard, S., Sandager, P., Petersen, O. B., Sørensen, T. S., Lundorf, E., and Stausbøl-Grøn, B. (2016). Overcoming foetal motion using interactive real-time magnetic resonance imaging. *Clinical Physiology and Functional Imaging*.
- [Brown et al., 2014] Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., Thompson, M. R., and Venkatesan, R. (2014). *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. John Wiley & Sons Ltd, Chichester, UK, second edition.
- [Buades et al., 2005] Buades, A., Coll, B., and Morel, J. M. (2005). A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling & Simulation*, 4(2):490–530.
- [Byrd et al., 1995] Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C. (1995). A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208.
- [Cachier et al., 2003] Cachier, P., Bardinet, E., Dormont, D., Pennec, X., and Ayache, N. (2003). Iconic feature based nonrigid registration: the PASHA algorithm. *Computer Vision and Image Understanding*, 89(2-3):272–298.

- [Candès and Fernandez-Granda, 2014] Candès, E. J. and Fernandez-Granda, C. (2014). Towards a Mathematical Theory of Super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956.
- [Candès et al., 2011] Candès, E. J., Li, X., Ma, Y., and Wright, J. (2011). Robust principal component analysis? *Journal of the ACM*, 58(3):1–37.
- [Cardoso et al., 2017] Cardoso, M. J., Arbel, T., Melbourne, A., Bogunovic, H., Moeskops, P., Chen, X., Schwartz, E., Garvin, M., Robinson, E., Trucco, E., **Ebner, M.**, Xu, Y., Makropoulos, A., Desjardin, A., and Vercauteren, T., editors (2017). *Fetal, Infant and Ophthalmic Medical Image Analysis*, volume 10554 of *Lecture Notes in Computer Science*. Springer International Publishing, Cham.
- [Cardoso et al., 2015] Cardoso, M. J., Modat, M., Vercauteren, T., and Ourselin, S. (2015). Scale Factor Point Spread Function Matching: Beyond Aliasing in Image Resampling. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 675–683. Springer International Publishing.
- [Carroll et al., 2007] Carroll, J. K., Herrick, B., Gipson, T., and Lee, S. P. (2007). Acute Pancreatitis: Diagnosis, Prognosis, and Treatment. *American Family Physician*, 75(10):1513–1520.
- [Casero et al., 2016] Casero, R., Siedlecka, U., Jones, E., Gruscheski, L., Johnston, C., Grau, V., and Kohl, P. (2016). Block-face imaging of wax-embedded tissue as a necessary step in the three-dimensional reconstruction of the heart structure. In *82th Annual Meeting of the German Cardiac Society–Cardiac and Circulation Research, March*, volume 30, page P1685.
- [Casero et al., 2017] Casero, R., Siedlecka, U., Jones, E. S., Gruscheski, L., Gibb, M., Schneider, J. E., Kohl, P., and Grau, V. (2017). Transformation diffusion reconstruction of three-dimensional histology volumes from two-dimensional image stacks. *Medical Image Analysis*, 38:184–204.
- [Chacko et al., 2015] Chacko, N., Chan, K. G., and Liebling, M. (2015). Intensity-based point-spread-function-aware registration for multi-view applications in op-

- tical microscopy. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 306–309. IEEE.
- [Chambolle and Pock, 2011] Chambolle, A. and Pock, T. (2011). A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145.
- [Charbonnier et al., 1997] Charbonnier, P., Blanc-Feraud, L., Aubert, G., and Barlaud, M. (1997). Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311.
- [Chathadi et al., 2015] Chathadi, K. V., Chandrasekhara, V., Acosta, R. D., Decker, G. A., Early, D. S., Eloubeidi, M. A., Evans, J. A., Faulx, A. L., Fanelli, R. D., Fisher, D. A., Foley, K., Fonkalsrud, L., Hwang, J. H., Jue, T. L., Khashab, M. A., Lightdale, J. R., Muthusamy, V. R., Pasha, S. F., Saltzman, J. R., Sharaf, R., Shaukat, A., Shergill, A. K., Wang, A., Cash, B. D., and DeWitt, J. M. (2015). The role of ERCP in benign diseases of the biliary tract. *Gastrointestinal Endoscopy*, 81(4):795–803.
- [Chen et al., 2015] Chen, W., Mo, J.-J., Lin, L., Li, C.-Q., and Zhang, J.-F. (2015). Diagnostic value of magnetic resonance cholangiopancreatography in choledocholithiasis. *World Journal of Gastroenterology*, 21(11):3351–3360.
- [Cheng et al., 2018] Cheng, X., Zhang, L., and Zheng, Y. (2018). Deep similarity learning for multimodal medical images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 6(3):248–252.
- [Chouhan et al., 2019] Chouhan, M., **Ebner, M.**, Patel, P. A., Atkinson, D., Firmin, L., Amin, Z., De Coppi, P., Ourselin, S., Vercauteren, T., and Taylor, S. A. (2019). Expert-reader validation of optimised Super-Resolution Reconstruction for upper abdominal MRI. In *European Society of Gastrointestinal and Abdominal Radiology (ESGAR)*.
- [Chung et al., 2019] Chung, K. K., Altmann, D., Barkhof, F., Miszkiel, K., Fisniku, L., Brex, P. A., O’Riordan, J., **Ebner, M.**, Prados, F., Cardoso, M. J., Ver-

- cauteren, T., Ourselin, S., Thompson, A., Ciccarelli, O., and Chard, D. T. (2019). A thirty year clinical and MRI observational study of multiple sclerosis and clinically isolated syndromes. *Under Revision (Annals of Neurology)*.
- [Çiçek et al., 2016] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G., and Wells, W., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 424–432, Cham. Springer International Publishing.
- [Cordero-Grande et al., 2018] Cordero-Grande, L., Hughes, E. J., Hutter, J., Price, A. N., and Hajnal, J. V. (2018). 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.
- [Coupé et al., 2008] Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., and Barillot, C. (2008). An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images. *IEEE Transactions on Medical Imaging*, 27(4):425–441.
- [Criminisi et al., 2009] Criminisi, A., Shotton, J., and Bucciarelli, S. (2009). Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes. *MICCAI Workshop on Probabilistic Models for Medical Image Analysis*, pages 69–80.
- [Dale et al., 2015] Dale, B. M., Brown, M. A., and Semelka, R. C. (2015). *MRI Basic Principles and Applications*. John Wiley & Sons, Ltd, Chichester, UK, 5th edition.
- [Damadian, 1971] Damadian, R. (1971). Tumor Detection by Nuclear Magnetic Resonance. *Science*, 171(3976):1151–1153.
- [De Stefano et al., 2014] De Stefano, N., Airas, L., Grigoriadis, N., Mattle, H. P., O’Riordan, J., Oreja-Guevara, C., Sellebjerg, F., Stankoff, B., Walczak, A.,

- Wiendl, H., and Kieseier, B. C. (2014). Clinical Relevance of Brain Volume Measures in Multiple Sclerosis. *CNS Drugs*, 28(2):147–156.
- [Diamond and Boyd, 2015] Diamond, S. and Boyd, S. (2015). Convex Optimization with Abstract Linear Operators. In *2015 IEEE International Conference on Computer Vision (ICCV)*, number 1, pages 675–683. IEEE.
- [Diwanji et al., 2016] Diwanji, N., Dibbad, R., Sastri, M., and Desai, E. (2016). The study of role of magnetic resonance cholangio-pancreatography (MRCP) in pancreato-biliary disorders. *International Journal of Medical Science and Public Health*, 5(12):2635.
- [Doel et al., 2017] Doel, T., Shakir, D. I., Pratt, R., Aertsen, M., Moggridge, J., Bellon, E., David, A. L., Deprest, J., Vercauteren, T., and Ourselin, S. (2017). GIFT-Cloud: A data sharing and collaboration platform for medical imaging research. *Computer Methods and Programs in Biomedicine*, 139:181–190.
- [Dong et al., 2016] Dong, C., Loy, C. C., He, K., and Tang, X. (2016). Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307.
- [Ebner et al., 2016] **Ebner, M.**, Aertsen, M., Pratt, R., David, A. L., Pandya, P. P., Deprest, J., Kendall, G., Klusmann, M., Humphries, P., Hewitt, R., Atkinson, D., Vercauteren, T., and Ourselin, S. (2016). Volumetric MRI Reconstruction of Trachea in Fetuses with Complex Fetal Neck Masses. In *Proceedings of the Institute of Physics and Engineering in Medicine (IPEM) Conference on Fetal, Neonatal and Paediatric MR Imaging: Techniques and Applications*, Leeds, United Kingdom.
- [Ebner et al., 2017a] **Ebner, M.**, Chouhan, M., Patel, P. A., Atkinson, D., Amin, Z., Read, S., Punwani, S., Taylor, S., Vercauteren, T., and Ourselin, S. (2017a). Point-Spread-Function-Aware Slice-to-Volume Registration: Application to Upper Abdominal MRI Super-Resolution. In Zuluaga, M. A., Bhatia, K., Kainz, B., Moghari, M. H., and Pace, D. F., editors, *Reconstruction, Segmentation, and*

- Analysis of Medical Images. RAMBO 2016*, volume 10129 of *Lecture Notes in Computer Science*, pages 3–13. Springer International Publishing.
- [Ebner et al., 2018a] **Ebner, M.**, Chung, K. K., Prados, F., Cardoso, M. J., Chard, D. T., Vercauteren, T., and Ourselin, S. (2018a). Volumetric reconstruction from printed films: Enabling 30 year longitudinal analysis in MR neuroimaging. *NeuroImage*, 165:238–250.
- [Ebner et al., 2018b] **Ebner, M.**, Modat, M., Ferraris, S., Ourselin, S., and Vercauteren, T. (2018b). Forward-backward splitting in deformable image registration: A demons approach. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1065–1069, Washington, D.C. IEEE.
- [Ebner et al., 2017b] **Ebner, M.**, Parker, K., Vercauteren, T., Ourselin, S., Wassertheurer, S., Hughes, A., and Hametner, B. (2017b). Calculating reservoir pressure with or without flow information: Similarity and algorithmic sensitivity at radial artery. *Artery Research*, 20:78–79.
- [Ebner et al., 2017c] **Ebner, M.**, Parker, K., Vercauteren, T., Ourselin, S., Wassertheurer, S., Hughes, A., and Hametner, B. (2017c). Reservoir pressure separation at brachial, carotid and radial arteries: A quantitative comparison and evaluation. *Artery Research*, 20:54.
- [Ebner et al., 2019a] **Ebner, M.**, Patel, P. A., Atkinson, D., Caselton, L., Firmin, L., Amin, Z., Bainbridge, A., De Coppi, P., Taylor, S. A., Ourselin, S., Chouhan, M. D., and Vercauteren, T. (2019a). Super-resolution for upper abdominal MRI: Acquisition and post-processing protocol optimization using brain MRI control data and expert reader validation. *Magnetic Resonance in Medicine*, page mrm.27852.
- [Ebner et al., 2019b] **Ebner, M.**, Patel, P. A., Atkinson, D., Caselton, L., Taylor, S. A., Bainbridge, A., Ourselin, S., Chouhan, M., and Vercauteren, T. (2019b). Reconstruction-based Super-Resolution for High-Resolution Abdominal MRI: A

- Preliminary Study. In *International Society for Magnetic Resonance in Medicine (ISMRM)*.
- [Ebner et al., 2018c] **Ebner, M.**, Wang, G., Li, W., Aertsen, M., Patel, P. A., Aughwane, R., Melbourne, A., Doel, T., David, A. L., Deprest, J., Ourselin, S., and Vercauteren, T. (2018c). An Automated Localization, Segmentation and Reconstruction Framework for Fetal Brain MRI. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2018*, pages 313–320. Springer.
- [Ebner et al., 2019c] **Ebner, M.**, Wang, G., Li, W., Aertsen, M., Patel, P. A., Aughwane, R., Melbourne, A., Doel, T., Dymarkowski, S., De Coppi, P., David, A. L., Deprest, J., Ourselin, S., and Vercauteren, T. (2019c). An Automated Framework for Localization, Segmentation and Super-Resolution Reconstruction of Fetal Brain MRI. *Under Submission (NeuroImage)*.
- [Ebner et al., 2018d] **Ebner, M.**, Xie, Y., Wykes, V., Miserocchi, A., McEvoy, A., Ourselin, S., and Vercauteren, T. (2018d). Spatial Regularisation based Reconstruction of Quantitative Fluorescence Imaging. In *Biophotonics Congress: Biomedical Optics Congress 2018 (Microscopy/Translational/Brain/OTS)*, page JTU3A.8, Hollywood, Florida United States. OSA.
- [Ekert et al., 2016] Ekert, K., Groeschel, S., Sánchez-Albisua, I., Frölich, S., Dieckmann, A., Engel, C., and Krägeloh-Mann, I. (2016). Brain morphometry in Pontocerebellar Hypoplasia type 2. *Orphanet Journal of Rare Diseases*, 11(1):100.
- [Engelbrecht et al., 2016] Engelbrecht, M., Rauws, E., and Besselink, M. (2016). The Value of Magnetic Resonance Cholangiopancreatography in Clinical Practice. *Journal of the American College of Surgeons*, 222(3):328.
- [Ethun et al., 2014] Ethun, C. G., Fallon, S. C., Cassady, C. I., Mehollin-Ray, A. R., Olutoye, O. O., Zamora, I. J., Lee, T. C., Welty, S. E., and Cass, D. L. (2014). Fetal MRI improves diagnostic accuracy in patients referred to a fetal center for suspected esophageal atresia. *Journal of Pediatric Surgery*, 49(5):712–715.

- [Ferrante and Paragios, 2017] Ferrante, E. and Paragios, N. (2017). Slice-to-volume medical image registration: A survey. *Medical Image Analysis*, 39:101–123.
- [Fidon et al., 2019] Fidon, L., **Ebner, M.**, Garcia-Peraza-Herrera, L. C., Modat, M., Ourselin, S., and Vercauteren, T. (2019). Incompressible image registration using divergence-conforming B-splines. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*.
- [Fisniku et al., 2008] Fisniku, L. K., Brex, P. A., Altmann, D. R., Miszkiel, K. A., Benton, C. E., Lanyon, R., Thompson, A. J., and Miller, D. H. (2008). Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis. *Brain*, 131(3):808–817.
- [Fogtmann et al., 2012] Fogtmann, M., Seshamani, S., Kim, K., Chapman, T., and Studholme, C. (2012). A unified approach for motion estimation and super resolution reconstruction from structural Magnetic Resonance imaging on moving subjects. In *MICCAI Workshop on Perinatal and Paediatric Imaging*, pages 9–16.
- [Fogtmann et al., 2014] Fogtmann, M., Seshamani, S., Kroenke, C., Cheng, X., Chapman, T., Wilm, J., Rousseau, F., and Studholme, C. (2014). A Unified Approach to Diffusion Direction Sensitive Slice Registration and 3-D DTI Reconstruction From Moving Fetal Brain Anatomy. *IEEE Transactions on Medical Imaging*, 33(2):272–289.
- [Garel, 2008] Garel, C. (2008). Imaging the fetus: when does MRI really help? *Pediatric Radiology*, 38(S3):467–470.
- [Gauriau et al., 2015] Gauriau, R., Cuingnet, R., Lesage, D., and Bloch, I. (2015). Multi-organ localization with cascaded global-to-local regression and shape prior. *Medical Image Analysis*, 23(1):70–83.
- [Gholipour et al., 2015] Gholipour, A., Afacan, O., Aganj, I., Scherrer, B., Prabhu, S. P., Sahin, M., and Warfield, S. K. (2015). Super-resolution reconstruction in frequency, image, and wavelet domains to reduce through-plane partial voluming in MRI. *Medical Physics*, 42(12):6919–6932.

- [Gholipour et al., 2014] Gholipour, A., Estroff, J. A., Barnewolt, C. E., Robertson, R. L., Grant, P. E., Gagoski, B., Warfield, S. K., Afacan, O., Connolly, S. A., Neil, J. J., Wolfberg, A., and Mulkern, R. V. (2014). Fetal MRI: A technical update with educational aspirations. *Concepts in Magnetic Resonance Part A*, 43(6):237–266.
- [Gholipour et al., 2010a] Gholipour, A., Estroff, J. a., Sahin, M., Prabhu, S. P., and Warfield, S. K. (2010a). Maximum A Posteriori Estimation of Isotropic High-Resolution Volumetric MRI from Orthogonal Thick-Slice Scans. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, volume 13, pages 109–116.
- [Gholipour et al., 2010b] Gholipour, A., Estroff, J. A., and Warfield, S. K. (2010b). Robust Super-Resolution Volume Reconstruction From Slice Acquisitions: Application to Fetal Brain MRI. *IEEE Transactions on Medical Imaging*, 29(10):1739–1758.
- [Gholipour et al., 2017] Gholipour, A., Rollins, C. K., Velasco-Annis, C., Oualam, A., Akhondi-Asl, A., Afacan, O., Ortinau, C. M., Clancy, S., Limperopoulos, C., Yang, E., Estroff, J. A., and Warfield, S. K. (2017). A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Scientific Reports*, 7(1):476.
- [Gholipour and Warfield, 2009] Gholipour, A. and Warfield, S. K. (2009). Super-resolution Reconstruction of Fetal Brain MRI. In *MICCAI Workshop on Image Analysis for the Developing Brain (IADB’2009)*. London, UK, pages 45–52.
- [Gibson et al., 2018] Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D. I., Wang, G., Eaton-Rosen, Z., Gray, R., Doel, T., Hu, Y., Whyntie, T., Nachev, P., Modat, M., Barratt, D. C., Ourselin, S., Cardoso, M. J., and Vercauteren, T. (2018). NiftyNet: a deep-learning platform for medical imaging. *Computer Methods and Programs in Biomedicine*, 158:113–122.
- [Goldstein et al., 2014] Goldstein, T., Studer, C., and Baraniuk, R. (2014). A

- Field Guide to Forward-Backward Splitting with a FASTA Implementation. *arXiv:1411.3406*.
- [Gracien et al., 2016] Gracien, R.-M., Reitz, S. C., Hof, S. M., Fleischer, V., Zimmermann, H., Droby, A., Steinmetz, H., Zipp, F., Deichmann, R., and Klein, J. C. (2016). Changes and variability of proton density and T1 relaxation times in early multiple sclerosis: MRI markers of neuronal damage in the cerebral cortex. *European Radiology*, 26(8):2578–2586.
- [Greenspan, 2008] Greenspan, H. (2008). Super-Resolution in Medical Imaging. *The Computer Journal*, 52(1):43–63.
- [Greenspan et al., 2002] Greenspan, H., Oz, G., Kiryati, N., and Peled, S. (2002). MRI inter-slice reconstruction using super-resolution. *Magnetic Resonance Imaging*, 20(5):437–446.
- [Griffin et al., 2012] Griffin, N., Charles-Edwards, G., and Grant, L. A. (2012). Magnetic resonance cholangiopancreatography: the ABC of MRCP. *Insights into Imaging*, 3(1):11–21.
- [Hansen, 1992] Hansen, P. C. (1992). Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. *SIAM Review*, 34(4):561–580.
- [Hansen, 2001] Hansen, P. C. (2001). The L-curve and its use in the numerical treatment of inverse problems. In *Computational Inverse Problems in Electrocardiology (Advances in Computational Bioengineering)*, volume 5, pages 119–142. WIT press Holland.
- [Hardie, 2010] Hardie, R. C. (2010). Super-Resolution Using Adaptive Wiener Filters. In Milanfar, P., editor, *Super-Resolution Imaging*, pages 35–61. CRC Press.
- [He et al., 2017] He, K., Gkioxari, G., Dollar, P., and Girshick, R. (2017). Mask R-CNN. In *The IEEE International Conference on Computer Vision (ICCV)*.
- [Hiasa et al., 2018] Hiasa, Y., Otake, Y., Takao, M., Matsuoka, T., Takashima, K., Prince, J. L., Sugano, N., and Sato, Y. (2018). Cross-modality image synthe-

- sis from unpaired data using CycleGAN: Effects of gradient consistency loss and training data size. pages 1–8.
- [Hou et al., 2017] Hou, B., Alansary, A., McDonagh, S., Davidson, A., Rutherford, M., Hajnal, J. V., Rueckert, D., Glocker, B., and Kainz, B. (2017). Predicting Slice-to-Volume Transformation in Presence of Arbitrary Subject Motion. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, pages 296–304. Springer International Publishing.
- [Hou et al., 2018] Hou, B., Khanal, B., Alansary, A., McDonagh, S., Davidson, A., Rutherford, M., Hajnal, J. V., Rueckert, D., Glocker, B., and Kainz, B. (2018). 3-D Reconstruction in Canonical Co-Ordinate Space From Arbitrarily Oriented 2-D Images. *IEEE Transactions on Medical Imaging*, 37(8):1737–1750.
- [Howard et al., 2006] Howard, K., Lord, S. J., Speer, A., Gibson, R. N., Padbury, R., and Kearney, B. (2006). Value of magnetic resonance cholangiopancreatography in the diagnosis of biliary abnormalities in postcholecystectomy patients: A probabilistic cost-effectiveness analysis of diagnostic strategies. *International Journal of Technology Assessment in Health Care*, 22(01):109–118.
- [Hu et al., 2018] Hu, Y., Modat, M., Gibson, E., Li, W., Ghavami, N., Bonmati, E., Wang, G., Bandula, S., Moore, C. M., Emberton, M., Ourselin, S., Noble, J. A., Barratt, D. C., and Vercauteren, T. (2018). Weakly-supervised convolutional neural networks for multimodal image registration. *Medical Image Analysis*, 49:1–13.
- [Isola et al., 2017] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976. IEEE.
- [Jenkinson et al., 2002] Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, 17(2):825–841.

- [Jiang et al., 2009] Jiang, S., Xue, H., Counsell, S., Anjari, M., Allsop, J., Rutherford, M., Rueckert, D., and Hajnal, J. V. (2009). Diffusion Tensor Imaging (DTI) of the Brain in Moving Subjects: Application to In-Utero Fetal and Ex-Utero Studies. *Magnetic Resonance in Medicine*, 62(3):645–655.
- [Jiang et al., 2006] Jiang, S., Xue, H., Glover, A., Rutherford, M., and Hajnal, J. V. (2006). A Novel Approach to Accurate 3D High Resolution and High SNR Fetal Brain Imaging. In *3rd IEEE International Symposium on Biomedical Imaging: Macro to Nano, 2006.*, volume 5, pages 662–665. IEEE.
- [Jiang et al., 2007] Jiang, S., Xue, H., Glover, A., Rutherford, M., Rueckert, D., and Hajnal, J. V. (2007). MRI of Moving Subjects Using Multislice Snapshot Images With Volume Reconstruction (SVR): Application to Fetal, Neonatal, and Adult Brain Studies. *IEEE Transactions on Medical Imaging*, 26(7):967–980.
- [Johnson et al., 2016] Johnson, J., Alahi, A., and Fei-Fei, L. (2016). Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9906 LNCS, pages 694–711.
- [Kainz et al., 2015a] Kainz, B., Alansary, A., Malamateniou, C., Keraudren, K., Rutherford, M., Hajnal, J. V., and Rueckert, D. (2015a). Flexible Reconstruction and Correction of Unpredictable Motion from Stacks of 2D Images. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 555–562. Springer International Publishing.
- [Kainz et al., 2014a] Kainz, B., Keraudren, K., Kyriakopoulou, V., Rutherford, M., Hajnal, J. V., and Rueckert, D. (2014a). Fast fully automatic brain detection in fetal MRI using dense rotation invariant image descriptors. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 1230–1233. IEEE.
- [Kainz et al., 2016] Kainz, B., Lloyd, D. F. A., Alansary, A., Murgasova, M. K., Khlebnikov, R., Rueckert, D., Rutherford, M., Razavi, R., and Hajnal, J. V. (2016). High-Performance Motion Correction of Fetal MRI. *EuroVis Workshop on*

- Reproducibility, Verification, and Validation in Visualization (EuroRV3)*, (July):3–5.
- [Kainz et al., 2014b] Kainz, B., Malamateniou, C., Murgasova, M., Keraudren, K., Rutherford, M., Hajnal, J. V., and Rueckert, D. (2014b). Motion Corrected 3D Reconstruction of the Fetal Thorax from Prenatal MRI. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, volume 17, pages 284–291.
- [Kainz et al., 2015b] Kainz, B., Steinberger, M., Wein, W., Kuklisova-Murgasova, M., Malamateniou, C., Keraudren, K., Torsney-Weir, T., Rutherford, M., Aljabar, P., Hajnal, J. V., and Rueckert, D. (2015b). Fast Volume Reconstruction From Motion Corrupted Stacks of 2D Slices. *IEEE Transactions on Medical Imaging*, 34(9):1901–1913.
- [Kanaan and Antaki, 2016] Kanaan, Z. and Antaki, F. (2016). Magnetic Resonance Cholangiopancreatography Still Plays a Role in the Preoperative Evaluation of Choledocholithiasis and Biliary Pathology. *Journal of the American College of Surgeons*, 222(3):325–326.
- [Keraudren et al., 2014] Keraudren, K., Kuklisova-Murgasova, M., Kyriakopoulou, V., Malamateniou, C., Rutherford, M., Kainz, B., Hajnal, J., and Rueckert, D. (2014). Automated fetal brain segmentation from 2D MRI slices for motion correction. *NeuroImage*, 101:633–643.
- [Keraudren et al., 2013] Keraudren, K., Kyriakopoulou, V., Rutherford, M., Hajnal, J. V., and Rueckert, D. (2013). Localisation of the Brain in Fetal MRI Using Bundled SIFT Features. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, volume 8149 LNCS, pages 582–589.
- [Kim et al., 1999] Kim, B., Boes, J. L., Bland, P. H., Chenevert, T. L., and Meyer, C. R. (1999). Motion Correction in fMRI via Registration of Individual Slices Into an Anatomical Volume. *Magnetic Resonance in Medicine*, 41(5):964–972.

- [Kim et al., 2010a] Kim, D., Sra, S., and Dhillon, I. S. (2010a). Tackling Box-Constrained Optimization via a New Projected Quasi-Newton Approach. *SIAM Journal on Scientific Computing*, 32(6):3548–3563.
- [Kim et al., 2013] Kim, D., Sra, S., and Dhillon, I. S. (2013). A non-monotonic method for large-scale non-negative least squares. *Optimization Methods and Software*, 28(5):1012–1039.
- [Kim et al., 2011a] Kim, K., Habas, P., Rajagopalan, V., Scott, J., Rousseau, F., Barkovich, A., Glenn, O., and Studholme, C. (2011a). Robust 3D Reconstruction from Motion Scattered Multislice MRI Using Second Order Models and Structure Tensor Weighted Kernel Regression. In *MICCAI Workshop on Image Analysis of Human Brain Development, Toronto, Canada*.
- [Kim et al., 2010b] Kim, K., Habas, P., Rousseau, F., Glenn, O., Barkovich, A., and Studholme, C. (2010b). Intersection Based Motion Correction of Multislice MRI for 3-D in Utero Fetal Brain Image Formation. *IEEE Transactions on Medical Imaging*, 29(1):146–158.
- [Kim et al., 2011b] Kim, K., Habas, P. A., Rajagopalan, V., Scott, J., Rousseau, F., Barkovich, A. J., Glenn, O. A., and Studholme, C. (2011b). SLIMMER: SLIce MRI motion estimation and reconstruction tool for studies of fetal anatomy. In Dawant, B. M. and Haynor, D. R., editors, *SPIE*, volume 7962, page 79624D.
- [Kim et al., 2011c] Kim, K., Habas, P. A., Rajagopalan, V., Scott, J. A., Corbett-Detig, J. M., Rousseau, F., Barkovich, a. J., Glenn, O. a., and Studholme, C. (2011c). Bias Field Inconsistency Correction of Motion-Scattered Multislice MRI for Improved 3D Image Reconstruction. *IEEE Transactions on Medical Imaging*, 30(9):1704–1712.
- [Kim et al., 2008] Kim, K., Hansen, M., Habas, P., Rousseau, F., Glenn, O., Barkovich, A. J., and Studholme, C. (2008). Intersection based registration of slice stacks to form 3D images of the human fetal brain. In *2008 5th IEEE Interna-*

- tional Symposium on Biomedical Imaging: From Nano to Macro*, pages 1167–1170. IEEE.
- [Kingma and Ba, 2015] Kingma, D. P. and Ba, J. L. (2015). Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*.
- [Kuklisova-Murgasova et al., 2012] Kuklisova-Murgasova, M., Quaghebeur, G., Rutherford, M. A., Hajnal, J. V., and Schnabel, J. A. (2012). Reconstruction of fetal brain MRI with intensity matching and complete outlier removal. *Medical Image Analysis*, 16(8):1550–1564.
- [Lawson and Hanson, 1974] Lawson, C. L. and Hanson, R. J. (1974). Linear least squares with linear inequality constraints. *Englewood Cliffs (NJ): Prentice-Hall*, 161.
- [Li et al., 2017a] Li, W., Wang, G., Fidon, L., Ourselin, S., Cardoso, M. J., and Vercauteren, T. (2017a). On the Compactness, Efficiency, and Representation of 3D Convolutional Networks: Brain Parcellation as a Pretext Task. volume 3565, pages 348–360.
- [Li et al., 2017b] Li, X., Guo, X., Ji, H., Yu, G., and Gao, P. (2017b). Gallstones in Patients with Chronic Liver Diseases. *BioMed Research International*, 2017:1–8.
- [Liang and Lauterbur, 2000] Liang, Z.-P. and Lauterbur, P. C. (2000). *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. Wiley-IEEE Press.
- [Lim et al., 2017] Lim, B., Son, S., Kim, H., Nah, S., and Lee, K. M. (2017). Enhanced Deep Residual Networks for Single Image Super-Resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140. IEEE.
- [Lin and Shum, 2004] Lin, Z. and Shum, H.-Y. (2004). Fundamental limits of

- reconstruction-based superresolution algorithms under local translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(1):83–97.
- [Lindeberg, 1994] Lindeberg, T. (1994). Scale-space theory: a basic tool for analyzing structures at different scales. *Journal of Applied Statistics*, 21(1-2):225–270.
- [Lloyd et al., 2019] Lloyd, D. F. A., Pushparajah, K., Simpson, J. M., van Amerom, J. F. P., van Poppel, M. P. M., Schulz, A., Kainz, B., Deprez, M., Lohezic, M., Allsop, J., Mathur, S., Bellsham-Revell, H., Vigneswaran, T., Charakida, M., Miller, O., Zidere, V., Sharland, G., Rutherford, M., Hajnal, J. V., and Razavi, R. (2019). Three-dimensional visualisation of the fetal heart using prenatal MRI with motion-corrected slice-volume registration: a prospective, single-centre cohort study. *The Lancet*, (18):1–10.
- [MacKay, 1998] MacKay, D. J. C. (1998). Introduction to Gaussian processes. *Neural Networks and Machine Learning*.
- [Makmun et al., 2017] Makmun, D., Fauzi, A., and Shatri, H. (2017). Sensitivity and Specificity of Magnetic Resonance Cholangiopancreatography versus Endoscopic Ultrasonography against Endoscopic Retrograde Cholangiopancreatography in Diagnosing Choledocholithiasis: The Indonesian Experience. *Clinical Endoscopy*, pages 1–5.
- [Malamateniou et al., 2013] Malamateniou, C., Malik, S. J., Counsell, S. J., Allsop, J. M., McGuinness, a. K., Hayat, T., Broadhouse, K., Nunes, R. G., Ederies, a. M., Hajnal, J. V., and Rutherford, M. a. (2013). Motion-Compensation Techniques in Neonatal and Fetal MR Imaging. *American Journal of Neuroradiology*, 34(6):1124–1136.
- [Manjón et al., 2010] Manjón, J. V., Coupé, P., Buades, A., Fonov, V., Louis Collins, D., and Robles, M. (2010). Non-local MRI upsampling. *Medical Image Analysis*, 14(6):784–792.
- [Mansi et al., 2011] Mansi, T., Pennec, X., Sermesant, M., Delingette, H., and Ayache, N. (2011). iLogDemons: A Demons-Based Registration Algorithm for Track-

- ing Incompressible Elastic Biological Tissues. *International Journal of Computer Vision*, 92(1):92–111.
- [Marami et al., 2017] Marami, B., Mohseni Salehi, S. S., Afacan, O., Scherrer, B., Rollins, C. K., Yang, E., Estroff, J. A., Warfield, S. K., and Gholipour, A. (2017). Temporal slice registration and robust diffusion-tensor reconstruction for improved fetal brain structural connectivity analysis. *NeuroImage*, 156:475–488.
- [Marami et al., 2016] Marami, B., Scherrer, B., Afacan, O., Erem, B., Warfield, S. K., and Gholipour, A. (2016). Motion-Robust Diffusion-Weighted Brain MRI Reconstruction Through Slice-Level Registration-Based Motion Tracking. *IEEE Transactions on Medical Imaging*, 35(10):2258–2269.
- [McClelland et al., 2013] McClelland, J. R., Hawkes, D. J., Schaeffter, T., and King, A. P. (2013). Respiratory motion models: A review. *Medical Image Analysis*, 17(1):19–42.
- [McCune et al., 1968] McCune, W. S., Shorrb, P. E., and Moscovitz, H. (1968). Endoscopic Cannulation of the Ampulla of Vater: A Preliminary Report. *Annals of Surgery*, 167(5):752–756.
- [McDonagh et al., 2017] McDonagh, S., Hou, B., Alansary, A., Oktay, O., Kamnitsas, K., Rutherford, M., Hajnal, J. V., and Kainz, B. (2017). Context-Sensitive Super-Resolution for Fast Fetal Magnetic Resonance Imaging. In *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment: Fifth International Workshop, CMMI 2017, Second International Workshop, RAMBO 2017, and First International Workshop, SWITCH 2017, Held in Conjunction*, pages 116—126. Springer International Publishing.
- [McPherson and Inder, 2017] McPherson, C. and Inder, T. (2017). Perinatal and neonatal use of sedation and analgesia. *Seminars in Fetal and Neonatal Medicine*, 22(5):314–320.
- [Melbourne et al., 2019] Melbourne, A., Wang, G., **Ebner, M.**, Vercauteren, T., Rosenbaum, D., Kasirer, Y., Wasertil, N., Ben-David, E., and Bin-Nun, A. (2019).

- T1 and T2 Weighted Image Segmentation from 1.0T Neonatal MRI. In *International Society for Magnetic Resonance in Medicine (ISMRM)*.
- [Miao et al., 2017] Miao, H., Mistelbauer, G., Karimov, A., Alansary, A., Davidson, A., Lloyd, D., Damodaram, M., Story, L., Hutter, J., Hajnal, J., Rutherford, M., Preim, B., Kainz, B., and Groller, M. E. (2017). Placenta Maps: In Utero Placental Health Assessment of the Human Fetus. *IEEE Transactions on Visualization and Computer Graphics*, 14(8):1–1.
- [Miao et al., 2016] Miao, S., Wang, Z. J., and Liao, R. (2016). A CNN Regression Approach for Real-Time 2D/3D Registration. *IEEE Transactions on Medical Imaging*, 35(5):1352–1363.
- [Milanfar, 2010] Milanfar, P. (2010). *Super-Resolution Imaging*. CRC Press.
- [Miller et al., 1988] Miller, D. H., Ormerod, I. E., McDonald, W. I., MacManus, D. G., Kendall, B. E., Kingsley, D. P., and Moseley, I. F. (1988). The early risk of multiple sclerosis after optic neuritis. *Journal of Neurology, Neurosurgery & Psychiatry*, 51(12):1569–1571.
- [Miller et al., 1989] Miller, D. H., Ormerod, I. E. C., Rudge, P., Kendall, B. E., Moseley, I. F., and McDonald, W. I. (1989). The early risk of multiple sclerosis following isolated acute syndromes of the brainstem and spinal cord. *Annals of Neurology*, 26(5):635–639.
- [Milletari et al., 2016] Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571. IEEE.
- [Modat et al., 2014] Modat, M., Cash, D. M., Daga, P., Winston, G. P., Duncan, J. S., and Ourselin, S. (2014). Global image registration using a symmetric block-matching approach. *Journal of Medical Imaging*, 1(2):024003–5.

- [Modat et al., 2010] Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C., and Ourselin, S. (2010). Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine*, 98(3):278–284.
- [Morales and Nocedal, 2011] Morales, J. L. and Nocedal, J. (2011). Remark on "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization". *ACM Transactions on Mathematical Software*, 38(1):1–4.
- [Morrissey et al., 1993] Morrissey, S. P., Miller, D. H., Kendall, B. E., Kingsley, D. P. E., Kelly, M. A., Francis, D. A., MacManus, D. G., and McDonald, W. I. (1993). The significance of brain magnetic resonance imaging abnormalities at presentation with clinically isolated syndromes suggestive of multiple sclerosis. *Brain*, 116(1):135–146.
- [Mugler, 2014] Mugler, J. P. (2014). Optimized three-dimensional fast-spin-echo MRI. *Journal of Magnetic Resonance Imaging*, 39(4):745–767.
- [Nielsen et al., 1996] Nielsen, M., Florack, L., and Deriche, R. (1996). Regularization, Scale-Space, and Edge Detection Filters. In *Computer Vision – ECCV 96*, pages 70–81. Springer Berlin Heidelberg.
- [Ochs et al., 2014] Ochs, P., Chen, Y., Brox, T., and Pock, T. (2014). iPiano: Inertial Proximal Algorithm for Non-Convex Optimization. *SIAM Journal on Imaging Sciences*, 7(2):1388–1419.
- [Odille et al., 2015] Odille, F., Bustin, A., Chen, B., Vuissoz, P.-A., and Felblinger, J. (2015). Motion-Corrected, Super-Resolution Reconstruction for High-Resolution 3D Cardiac Cine MRI. In Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F., editors, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 9351 of *Lecture Notes in Computer Science*, pages 435–442. Springer International Publishing, Cham.

- [Okta et al., 2016] Oktay, O., Bai, W., Lee, M., Guerrero, R., Kamnitsas, K., Caballero, J., de Marvao, A., Cook, S., O'Regan, D., and Rueckert, D. (2016). Multi-input Cardiac Image Super-Resolution Using Convolutional Neural Networks. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 246–254. Springer International Publishing, Athens, Greece.
- [Okta et al., 2017] Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S., de Marvao, A., Dawes, T., O'Regan, D., Kainz, B., Glocker, B., and Rueckert, D. (2017). Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation. *IEEE Transactions on Medical Imaging*, pages 1–1.
- [Okta et al., 2018] Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D. (2018). Attention U-Net: Learning Where to Look for the Pancreas. In *Medical Imaging with Deep Learning – MIDL 2018*.
- [O'Riordan et al., 1998] O'Riordan, J. I., Thompson, A. J., Kingsley, D. P. E., MacManus, D. G., Kendall, B. E., Rudge, P., McDonald, W. I., and Miller, D. H. (1998). The prognostic value of brain MRI in clinically isolated syndromes of the CNS. A 10-year follow-up. *Brain*, 121(3):495–503.
- [Ormerod et al., 1987] Ormerod, I. E. C., Miller, D. H., McDonald, W. I., Du Boulay, E. P. G. H., Rudge, P., Kendall, B. E., Moseley, I. F., Johnson, G., Tofts, P. S., Halliday, A. M., Bronstein, A. M., Scaravilli, F., Harding, A. E., Barnes, D., and Zilkha, K. J. (1987). The role of NMR imaging in the assessment of multiple sclerosis and isolated neurological lesions. *Brain*, 110(6):1579–1616.
- [Ovaere et al., 2015] Ovaere, C., Eggink, A., Richter, J., Cohen-Overbeek, T. E., Van Calenbergh, F., Jansen, K., Oepkes, D., Devlieger, R., De Catte, L., and Deprest, J. A. (2015). Prenatal Diagnosis and Patient Preferences in Patients with Neural Tube Defects around the Advent of Fetal Surgery in Belgium and Holland. *Fetal Diagnosis and Therapy*, 37(3):226–234.

- [Papież et al., 2014] Papież, B. W., Heinrich, M. P., Fehrenbach, J., Risser, L., and Schnabel, J. A. (2014). An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. *Medical Image Analysis*, 18(8):1299–1311.
- [Parikh and Boyd, 2014] Parikh, N. and Boyd, S. (2014). Proximal Algorithms. *Foundations and Trends® in Optimization*, 1(3):127–239.
- [Park et al., 2003] Park, S. C., Park, M. K., and Kang, M. G. (2003). Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36.
- [Parthasarathy et al., 2016] Parthasarathy, M., Maqsood, H., Sill, A. M., Kowdley, G. C., and Cunningham, S. C. (2016). Abandoning Hasty Conclusions: The Use of Magnetic Resonance Cholangiopancreatography in Clinical Practice. *Journal of the American College of Surgeons*, 222(3):326–328.
- [Peled and Yeshurun, 2002] Peled, S. and Yeshurun, Y. (2002). Superresolution in MRI – Perhaps sometimes. *Magnetic Resonance in Medicine*, 48(2):409–409.
- [Pennec et al., 1999] Pennec, X., Cachier, P., and Ayache, N. (1999). Understanding the “ demons ” Algorithm: 3D Non-rigid Registration by Gradient Descent. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 1999*, pages 597–605.
- [Plenge et al., 2012] Plenge, E., Poot, D. H. J., Bernsen, M., Kotek, G., Houston, G., Wielopolski, P., van der Weerd, L., Niessen, W. J., and Meijering, E. (2012). Super-resolution methods in MRI: Can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? *Magnetic Resonance in Medicine*, 68(6):1983–1993.
- [Pock et al., 2007] Pock, T., Urschler, M., Zach, C., Beichel, R., and Bischof, H. (2007). A Duality Based Algorithm for TV-L1-Optical-Flow Image Registration. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007*, volume 10, pages 511–518. Springer Berlin Heidelberg, Berlin, Heidelberg.

- [Poot et al., 2010] Poot, D. H. J., Van Meir, V., and Sijbers, J. (2010). General and Efficient Super-Resolution Method for Multi-slice MRI. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, volume 6361 LNCS, pages 615–622. Springer Berlin Heidelberg.
- [Prayer et al., 2004] Prayer, D., Brugger, P., and Prayer, L. (2004). Fetal MRI: techniques and protocols. *Pediatric Radiology*, 34(9):685–693.
- [Protter and Elad, 2010] Protter, M. and Elad, M. (2010). Super-Resolution With Probabilistic Motion Estimation. In Milanfar, P., editor, *Super-Resolution Imaging*, pages 97–121. CRC Press.
- [Protter et al., 2009] Protter, M., Elad, M., Takeda, H., and Milanfar, P. (2009). Generalizing the Nonlocal-Means to Super-Resolution Reconstruction. *IEEE Transactions on Image Processing*, 18(1):36–51.
- [Purcell et al., 1946] Purcell, E. M., Torrey, H. C., and Pound, R. V. (1946). Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review*, 69(1-2):37–38.
- [Rajchl et al., 2016] Rajchl, M., Lee, M. C. H., Schrans, F., Davidson, A., Passerat-Palmbach, J., Tarroni, G., Alansary, A., Oktay, O., Kainz, B., and Rueckert, D. (2016). Learning under distributed weak supervision. *arXiv preprint arXiv:1606.01100*.
- [Ranzini et al., 2018] Ranzini, M. B. M., **Ebner, M.**, Cardoso, M. J., Fotiadou, A., Vercauteren, T., Henckel, J., Hart, A., Ourselin, S., and Modat, M. (2018). Joint Multimodal Segmentation of Clinical CT and MR from Hip Arthroplasty Patients. In *MICCAI Workshop on Computational Methods and Clinical Applications in Musculoskeletal Imaging (MSKI) 2017*, volume 10734 LNCS, pages 72–84. Springer.
- [Ranzini et al., 2019] Ranzini, M. B. M., Henckel, J., **Ebner, M.**, Cardoso, M. J., Vercauteren, T., Ourselin, S., Hart, A., and Modat, M. (2019). Hip arthroplasty

- failure assessment using multimodal imaging joint segmentation. *Under Revision (Computer Methods and Programs in Biomedicine)*.
- [Ringe and Wacker, 2015] Ringe, K. I. and Wacker, F. (2015). Radiological diagnosis in cholangiocarcinoma: Application of computed tomography, magnetic resonance imaging, and positron emission tomography. *Best Practice & Research Clinical Gastroenterology*, 29(2):253–265.
- [Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241.
- [Rousseau et al., 2005] Rousseau, F., Glenn, O., Iordanova, B., Rodriguez-Carranza, C., Vigneron, D., Barkovich, J., and Studholme, C. (2005). A Novel Approach to High Resolution Fetal Brain MR Imaging. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, volume 8, pages 548–555. Springer.
- [Rousseau et al., 2006] Rousseau, F., Glenn, O. A., Iordanova, B., Rodriguez-Carranza, C., Vigneron, D. B., Barkovich, J. A., and Studholme, C. (2006). Registration-Based Approach for Reconstruction of High-Resolution In Utero Fetal MR Brain Images. *Academic Radiology*, 13(9):1072–1081.
- [Rousseau et al., 2010] Rousseau, F., Kim, K., Studholme, C., Koob, M., and Dietemann, J. L. (2010). On Super-Resolution for Fetal Brain MRI. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*, pages 355–362.
- [Rousseau et al., 2013] Rousseau, F., Oubel, E., Pontabry, J., Schweitzer, M., Studholme, C., Koob, M., and Dietemann, J.-L. (2013). BTK: An open-source toolkit for fetal brain MR image processing. *Computer Methods and Programs in Biomedicine*, 109(1):65–73.
- [Rudin et al., 1992] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total

- variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268.
- [Sacco et al., 2018] Sacco, A., Simpson, L., Deprest, J., and David, A. L. (2018). A study to assess global availability of fetal surgery for myelomeningocele. *Prenatal Diagnosis*, 38(13):1020–1027.
- [Safari et al., 2016] Safari, M. T., Miri, M. B., Ebadi, S., Shahrokh, S., and Alizadeh, A. H. M. (2016). Comparing the Roles of EUS, ERCP and MRCP in Idiopathic Acute Recurrent Pancreatitis. *Clinical Medicine Insights: Gastroenterology*, 9:35–39.
- [Sailer et al., 1999] Sailer, M., O’Riordan, J. I., Thompson, A. J., Kingsley, D. P., MacManus, D. G., McDonald, W. I., and Miller, D. H. (1999). Quantitative MRI in patients with clinically isolated syndromes suggestive of demyelination. *Neurology*, 52(3):599–599.
- [Saleem, 2014] Saleem, S. N. (2014). Fetal MRI: An approach to practice: A review. *Journal of Advanced Research*, 5(5):507–523.
- [Salehi et al., 2018] Salehi, S. S. M., Hashemi, S. R., Velasco-Annis, C., Oualam, A., Estroff, J. A., Erdogmus, D., Warfield, S. K., and Gholipour, A. (2018). Real-time automatic fetal brain extraction in fetal MRI by deep learning. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 720–724.
- [Salvatier et al., 2016] Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55.
- [Santos-Ribeiro et al., 2016] Santos-Ribeiro, A., Nutt, D. J., and McGonigle, J. (2016). Inertial Demons: A Momentum-Based Diffeomorphic Registration Framework. In Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G., and Wells, W., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science*, volume 9902 of *Lecture Notes in Computer Science*, pages 37–45. Springer International Publishing, Cham.

- [Scheffler, 2002] Scheffler, K. (2002). Superresolution in MRI? *Magnetic Resonance in Medicine*, 48(2):408–408.
- [Scott and Jewell, 2019] Scott, E. and Jewell, A. (2019). Supportive care needs of people with pancreatic cancer: a literature review. *Cancer Nursing Practice*.
- [Sepulveda et al., 2012] Sepulveda, W., Wong, A. E., Sepulveda, F., Martinez-Ten, P., and Ximenes, R. (2012). Fetal magnetic resonance imaging and three-dimensional ultrasound in clinical practice: General aspects. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 26(5):575–591.
- [Seshamani et al., 2016] Seshamani, S., Blazejewska, A. I., Mckown, S., Caucutt, J., Dighe, M., Gatenby, C., and Studholme, C. (2016). Detecting Default Mode Networks In Utero by Integrated 4D fMRI Reconstruction and Analysis. *Human Brain Mapping*, 37(11):4158–4178.
- [Shi et al., 2016] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. (2016). Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1874–1883.
- [Shi et al., 2013] Shi, W., Caballero, J., Ledig, C., Zhuang, X., Bai, W., Bhatia, K., de Marvao, A. M. S. M., Dawes, T., O’Regan, D., and Rueckert, D. (2013). Cardiac Image Super-Resolution with Global Correspondence Using Multi-Atlas PatchMatch. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, volume 8151, pages 9–16. Springer Berlin Heidelberg.
- [Shiani et al., 2018] Shiani, A., Lipka, S., Wolk, B., Pinkas, H., Kumar, A., Alsina, A., Kemmer, N., Turner, A., and Brady, P. (2018). Diagnostic Accuracy of Magnetic Resonance Cholangiopancreatography Versus Endoscopic Retrograde Cholangiopancreatography Findings in the Postorthotopic Liver Transplant Population. *European Medical Journal*, 3(3):58–64.

- [Shilling et al., 2009] Shilling, R., Robbie, T., Bailloeu, T., Mewes, K., Mersereau, R., and Brummer, M. (2009). A Super-Resolution Framework for 3-D High-Resolution and High-Contrast Imaging Using 2-D Multislice MRI. *IEEE Transactions on Medical Imaging*, 28(5):633–644.
- [Siedlecka et al., 2013] Siedlecka, U., Casero, R., Burton, R. A., Afonso, C., Bolensdorff, C., Grau, V., and Kohl, P. (2013). Three-dimensional histological reconstruction of the heart guided by block-face imaging of wax-embedded tissue. In *6th International Workshop on Cardiac Mechano-Electric Coupling and Arrhythmias*, volume 12.
- [Siegel et al., 2019] Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, 69(1):7–34.
- [Simonovsky et al., 2016] Simonovsky, M., Gutiérrez-Becker, B., Mateus, D., Navab, N., and Komodakis, N. (2016). A Deep Metric for Multimodal Registration. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, pages 10–18. Springer, Cham.
- [Smith, 2002] Smith, S. M. (2002). Fast Robust Automated Brain Extraction. *Human Brain Mapping*, 17(3):143–155.
- [Smith et al., 2002] Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P., Federico, A., and De Stefano, N. (2002). Accurate, Robust, and Automated Longitudinal and Cross-Sectional Brain Change Analysis. *NeuroImage*, 17(1):479–489.
- [Sorel et al., 2010] Sorel, M., Sroubek, F., and Flusser, J. (2010). Towards Super-Resolution in the Presence of Spatially Varying Blur. In Milanfar, P., editor, *Super-Resolution Imaging*, pages 187–217. CRC Press.
- [Story et al., 2017] Story, L., Zhan, T., Aljabar, P., Hajnal, J. V., Shennan, A., and Rutherford, M. A. (2017). Magnetic resonance imaging assessment of lung volumes in fetuses at high risk of preterm birth. *BJOG: An International Journal of Obstetrics & Gynaecology: Fetal Medicine Poster Abstracts*, 124:17–49.

- [Sudre et al., 2017] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M. (2017). Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support – DLMIA 2017*, volume 10553, pages 240–248.
- [Takeda et al., 2009] Takeda, H., Milanfar, P., Protter, M., and Elad, M. (2009). Super-Resolution Without Explicit Subpixel Motion Estimation. *IEEE Transactions on Image Processing*, 18(9):1958–1975.
- [Taleb et al., 2013] Taleb, Y., Schweitzer, M., Studholme, C., Koob, M. M., Dietemann, J.-L., and Rousseau, F. F. (2013). Automatic template-based brain extraction in fetal MR images. *OHBM*.
- [Tanaka and Okutomi, 2010] Tanaka, M. and Okutomi, M. (2010). Toward Robust Reconstruction-Based Super-Resolution. In Milanfar, P., editor, *Super-Resolution Imaging*, pages 219–246. CRC Press.
- [Tanner et al., 2018] Tanner, C., Ozdemir, F., Profanter, R., Vishnevsky, V., Konukoglu, E., and Goksel, O. (2018). Generative Adversarial Networks for MR-CT Deformable Image Registration. pages 1–11.
- [Thirion, 1998] Thirion, J.-P. (1998). Image matching as a diffusion process: an analogy with Maxwell’s demons. *Medical Image Analysis*, 2(3):243–260.
- [Tourbier et al., 2014] Tourbier, S., Bresson, X., Hagmann, P., Thiran, J.-P., Meuli, R., and Cuadra, M. B. (2014). Efficient Total Variation Algorithm for Fetal Brain MRI Reconstruction. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, volume 17, pages 252–259.
- [Tourbier et al., 2015] Tourbier, S., Bresson, X., Hagmann, P., Thiran, J.-P., Meuli, R., and Cuadra, M. B. (2015). An efficient total variation algorithm for super-resolution in fetal brain MRI with adaptive regularization. *NeuroImage*, 118:1–14.

- [Tourbier et al., 2017] Tourbier, S., Velasco-Annis, C., Taimouri, V., Hagmann, P., Meuli, R., Warfield, S. K., Bach Cuadra, M., and Gholipour, A. (2017). Automated template-based brain localization and extraction for fetal brain MRI reconstruction. *NeuroImage*, 155:460–472.
- [Tustison et al., 2010] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320.
- [Unser et al., 1993] Unser, M., Aldroubi, A., and Eden, M. (1993). B-Spline Signal Processing: Part I – Theory. *IEEE Transactions on Signal Processing*, 41(2):821–833.
- [van Dijk et al., 2017] van Dijk, A. H., de Reuver, P. R., Besselink, M. G., van Laarhoven, K. J., Harrison, E. M., Wigmore, S. J., Hugh, T. J., and Boermeester, M. A. (2017). Assessment of available evidence in the management of gallbladder and bile duct stones: a systematic review of international guidelines. *HPB*, 19(4):297–309.
- [Van Reeth et al., 2015] Van Reeth, E., Tan, C. H., Tham, I. W., and Poh, C. L. (2015). Isotropic Reconstruction of a 4-D MRI Thoracic Sequence Using Super-Resolution. *Magnetic Resonance in Medicine*, 73(2):784–793.
- [Van Reeth et al., 2012] Van Reeth, E., Tham, I. W. K., Tan, C. H., and Poh, C. L. (2012). Super-resolution in magnetic resonance imaging: A review. *Concepts in Magnetic Resonance Part A*, 40A(6):306–325.
- [Vandewalle et al., 2010] Vandewalle, P., Sbaiz, L., and Vetterli, M. (2010). Registration for Super-Resolution: Theory, Algorithms, and Applications in Image and Mobile Video Enhancement. In Milanfar, P., editor, *Super-Resolution Imaging*, pages 155–185. CRC Press.
- [Vercauteren et al., 2007] Vercauteren, T., Pennec, X., Malis, E., Perchant, A., and Ayache, N. (2007). Insight into Efficient Image Registration Techniques and the

- Demons Algorithm. In *Information Processing in Medical Imaging*, volume 20, pages 495–506. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Vercauteren et al., 2008] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2008). Symmetric Log-Domain Diffeomorphic Registration: A Demons-Based Approach. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*, pages 754–761.
- [Vercauteren et al., 2009] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72.
- [Vercauteren et al., 2006] Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., and Ayache, N. (2006). Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. *Medical Image Analysis*, 10(5):673–692.
- [Vidal-Migallon et al., 2013] Vidal-Migallon, I., Commowick, O., Pennec, X., Daudet, J., and Vercauteren, T. (2013). GPU and CPU implementation of Young - Van Vliet’s Recursive Gaussian Smoothing Filter. *Insight Journal (ITK)*.
- [Villard et al., 2017] Villard, B., Zacur, E., Dall’Armellina, E., and Grau, V. (2017). Correction of Slice Misalignment in Multi-breath-hold Cardiac MRI Scans. volume 10124 of *Lecture Notes in Computer Science*, pages 30–38. Springer International Publishing, Cham.
- [Wallner et al., 1991] Wallner, B. K., Schumacher, K. A., Weidenmaier, W., and Friedrich, J. M. (1991). Dilated Biliary Tract: Evaluation with MR Cholangiography with a T2-weighted Contrast-enhanced Fast Sequence. *Radiology*, 181(3):805–808.
- [Walshe et al., 2016] Walshe, T. M., McLean, K. A., Patel, R., Chang, S. D., and Harris, A. C. (2016). Imaging of the biliary tree: Infection, inflammation and infiltratio. *Applied Radiology*, 45(4):20–26.

- [Wang et al., 2018] Wang, G., Li, W., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M., Doel, T., David, A. L., Deprest, J., Ourselin, S., and Vercauteren, T. (2018). Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning. *IEEE Transactions on Medical Imaging*, 37(7):1562–1573.
- [Wang et al., 2004] Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612.
- [Williams et al., 1999] Williams, D. B., Sahai, A. V., Aabakken, L., Penman, I. D., van Velse, A., Webb, J., Wilson, M., Hoffman, B. J., and Hawes, R. H. (1999). Endoscopic ultrasound guided fine needle aspiration biopsy: a large single centre experience. *Gut*, 44(5):720–726.
- [Wolterink et al., 2017] Wolterink, J. M., Dinkla, A. M., Savenije, M. H. F., Seevinck, P. R., van den Berg, C. A. T., and Isgum, I. (2017). Deep MR to CT Synthesis Using Unpaired Data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10557 LNCS, pages 14–23.
- [Woo et al., 2012] Woo, J., Murano, E. Z., Stone, M., and Prince, J. L. (2012). Reconstruction of High-Resolution Tongue Volumes From MRI. *IEEE Transactions on Biomedical Engineering*, 59(12):3511–3524.
- [Xie et al., 2017] Xie, Y., Thom, M., **Ebner, M.**, Wykes, V., Desjardins, A., Mierocchi, A., Ourselin, S., McEvoy, A. W., and Vercauteren, T. (2017). Wide-field spectrally resolved quantitative fluorescence imaging system: toward neurosurgical guidance in glioma resection. *Journal of Biomedical Optics*, 22(11):1–14.
- [Yang et al., 2019] Yang, D.-J., Xiong, J.-J., Lu, H.-M., Wei, Y., Zhang, L., Lu, S., and Hu, W.-M. (2019). The oncological safety in minimally invasive versus open distal pancreatectomy for pancreatic ductal adenocarcinoma: a systematic review and meta-analysis. *Scientific Reports*, 9(1):1159.

- [Yang and Huang, 2010] Yang, J. and Huang, T. (2010). Image super-resolution: Historical overview and future challenges. In Milanfar, P., editor, *Super-Resolution Imaging*, pages 20–34. CRC Press.
- [Yu and Koltun, 2016] Yu, F. and Koltun, V. (2016). Multi-Scale Context Aggregation by Dilated Convolutions. In *International Conference on Learning Representations – ICLR 2016*.
- [Zhu et al., 2017] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, volume 2017-Octob, pages 2242–2251. IEEE.