Correcting Susceptibility Artifacts in High-resolution Brain Magnetic Resonance Imaging

Soan Thi Minh Duong

Follow this and additional works at: https://ro.uow.edu.au/theses1
Correcting Susceptibility Artifacts in High-resolution Brain Magnetic Resonance Imaging

Soan Thi Minh Duong

This thesis is presented as part of the requirements for the conferral of the degree:

Doctor of Philosophy

Supervisor:
Assoc. Prof. Son Lam Phung

Co-supervisors:
Dr. Mark Schira & Prof. Abdesselam Bouzerdoum

The University of Wollongong
School of Electrical, Computer and Telecommunications Engineering

September 2020
This work © copyright by Soan Thi Minh Duong, 2020. All Rights Reserved.

No part of this work may be reproduced, stored in a retrieval system, transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of the author or the University of Wollongong.

This research has been conducted with the support of an Australian Government Research Training Program Scholarship.
Declaration

I, Soan Thi Minh Duong, declare that this thesis is submitted in partial fulfilment of the requirements for the conferral of the degree Doctor of Philosophy, from the University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Soan Thi Minh Duong

September 17, 2020
Abstract

Echo planar imaging (EPI) is a fast and non-invasive magnetic resonance imaging (MRI) technique that supports data acquisition at high spatial and temporal resolutions. Thus, EPI is widely used for human brain studies in both clinical diagnosis and scientific investigation. However, susceptibility artifacts, which cause the misalignment to the underlying structural image, are unavoidable distortions in EPI. These distortions are especially severe in high spatial-resolution images and can lead to misrepresentation of the human brain functions. Many susceptibility artifact correction (SAC) methods have been developed to address these challenges, but they require high computational resources, modified scanner hardware, or a modified acquisition protocol.

This thesis investigates existing SAC methods and develops new alternatives for high-resolution brain EPI images. The aims of developing new SAC methods are to reduce the computational cost and improve the correction accuracy. Three novel SAC methods, which are from two main categories: traditional iterative-optimization and deep learning, are proposed and tested in this research.

We propose a new iterative-optimization method using a T1 weighted (T1w) image to correct the susceptibility artifacts in a pair of reversed phase-encoding images. The reversed phase-encoding image pair is two images acquired using the same sequence but with inverse phase-encoding directions. The structural information from the T1w image is used to regularize the correction, and to select
the regularization parameters automatically. The proposed iterative-optimization method achieves more robust and sharper corrections and runs faster in comparison with other state-of-the-art SAC methods.

We develop a new unsupervised deep learning network for correcting the susceptibility artifacts in a pair of reversed phase-encoding images. To the best of our knowledge, the proposed method is the first attempt at using a learning-based approach with a deep network to correct the susceptibility artifacts. The proposed network is trained in an unsupervised manner, without the use of ground-truth information. For a new reversed phase-encoding image pair, corrected images are obtained directly by evaluating the trained model. The experimental results show that the proposed method produces the corrected images with the same accuracy as the state-of-the-art SAC methods while running many times faster (in second). These results indicate the feasibility of using the deep learning approach to correct the susceptibility artifacts.

We extend the deep learning SAC model by adding size and value normalization layers, and utilizing $T_{1w}$ images in the training phase. These extensions aim to make the network more robust to the size, resolution, and modality of the input images. The experimental results show that the improved deep network produces corrected images with better accuracy than the state-of-the-art SAC methods while running dramatically faster. Furthermore, the proposed network is able to transfer the learned features from one dataset to another dataset, thereby reducing the training time needed for a new dataset.
Acknowledgments

I would like to express my sincerest thanks to my supervisors, A/Prof. Son Lam Phung, Dr Mark Schira, and Prof. Abdesselam Bouzerdoum for all their counsel, encouragement, inspiration, and knowledge. I especially thank Lam for providing me with his valuable time and technical support. His calm and easy manner of teaching and patience has gone such a long way towards giving the skills that I need not only to complete the PhD project but also to pursue the research career in the future. I also thank Lam and Mark for a number of great life lessons that they imparted me during years of study.

I gratefully acknowledge the UOW Security team without whom I could not go home safety at late hours.

I would like to thank my fellow labmates, Anh, Thanh, Xiaolin, Paul, Zoey, and Harriet. We had many fruitful discussions regarding my research that helped me conceive new ideas. The recreational activities that we spent together were so productive and made my PhD life colorful.

I am thankful to my housemates and friends, who helped and took care of me in every matter during my stay in Wollongong. I acknowledge Trang, Ly, Cuong, Dr Trung, Truc, and friends who lived in 30 Cabbage, 21 Crown, 19 Hindmarsh, and 30 Market houses. We shared many delightful and memorable moments.

And finally, I would like to express my gratitude to my Mum, Dad, little sister, and my big family, who have continuously supported and encouraged me during my PhD journey. Your endless loves have helped me overcome life obstacles.
Contents

Abstract iv

Acknowledgments vi

List of figures x

List of tables xvi

Acronyms xviii

1 Introduction 1

1.1 Research motivation and objectives . . . . . . . . . . . . . . . . . . 1

1.2 Thesis organization . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

1.3 Research contributions . . . . . . . . . . . . . . . . . . . . . . . . . 4

1.4 Publications . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

2 Background and Related Work 7

2.1 Fundamentals of susceptibility artifacts in MRI . . . . . . . . . . . . 8

2.1.1 Susceptibility artifacts in conventional 2D GRE imaging . . . . 10

2.1.2 Susceptibility artifacts in EPI . . . . . . . . . . . . . . . . . . . 13

2.2 Existing susceptibility artifact correction methods . . . . . . . . . . 15

2.2.1 Fieldmap-based SACs . . . . . . . . . . . . . . . . . . . . . . . 15

2.2.2 Point spread function based SACs . . . . . . . . . . . . . . . . 17

2.2.3 Reversed-PE based SACs . . . . . . . . . . . . . . . . . . . . . 17
## Contents

2.2.4 Image registration based SACs ........................................... 18  
2.2.5 Hybrid SACs ............................................................... 19  
2.3 Chapter summary ............................................................ 20  

3 Susceptibility Artifact Correction using T1-weighted Regularization 21  
3.1 Introduction ................................................................. 22  
3.2 Hyper-elastic susceptibility artefact correction .............................. 23  
3.3 $T_{1w}$-guided reversed-PE SAC ........................................... 27  
    3.3.1 TISAC registration ..................................................... 27  
    3.3.2 Optimization of hyper-parameters .................................. 30  
3.4 Experiments and results ..................................................... 32  
    3.4.1 Data acquisition and preprocessing ................................ 33  
    3.4.2 Performance measures .............................................. 36  
    3.4.3 Experimental methods .............................................. 38  
    3.4.4 Analysis of the proposed method .................................. 40  
    3.4.5 Comparisons of SAC methods .................................... 40  
    3.4.6 Discussion ............................................................ 45  
3.5 Chapter summary ............................................................ 46  

4 Susceptibility Artifact Correction with Unsupervised Deep Learning 47  
4.1 Introduction ................................................................. 48  
4.2 Related work ............................................................... 49  
4.3 Unsupervised deep learning framework for SAC ............................ 51  
    4.3.1 CNN architecture for mapping ...................................... 53  
    4.3.2 Spatial transformation unit for unwarping ....................... 54  
    4.3.3 Loss function for self-supervised network learning .......... 55  
4.4 Experiments and analysis .................................................. 57  
    4.4.1 Brain EPI datasets .................................................... 57  
    4.4.2 Experimental methods .............................................. 59  
    4.4.3 Analysis of hyper-parameters ..................................... 60  

viii
4.4.4 Comparison with other methods in correction accuracy . . . 61
4.4.5 Comparison with other methods in processing speed . . . . 65
4.4.6 Discussion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67
4.5 Chapter summary . . . . . . . . . . . . . . . . . . . . . . . . . . . 68

5 Susceptibility Artifact Correction with Multi-modal Deep Learning 69
5.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 69
5.2 Deep learning framework for SAC . . . . . . . . . . . . . . . . . 71
5.2.1 Deep network for mapping . . . . . . . . . . . . . . . . . . . 72
5.2.2 Loss function for TS-Net training . . . . . . . . . . . . . . . . 73
5.3 Experiments and analysis . . . . . . . . . . . . . . . . . . . . . . . 76
5.3.1 Image datasets and experimental methods . . . . . . . . . . . 76
5.3.2 Ablation study of the proposed method . . . . . . . . . . . . . 78
5.3.3 Comparison with other methods . . . . . . . . . . . . . . . . . 82
5.4 Chapter summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . 87

6 Conclusion 89
6.1 Research summary . . . . . . . . . . . . . . . . . . . . . . . . . . 90
6.2 Future work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 91
6.3 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 92

A Principles of Magnetic Resonance Imaging 94
A.1 MR signal generation . . . . . . . . . . . . . . . . . . . . . . . . . 94
A.2 MR image formation . . . . . . . . . . . . . . . . . . . . . . . . . 99

B Multi-modal similarity measures 102

References 105
List of Figures

1.1 Example of the susceptibility artifacts in a pair of LR and RL phase encoding EPI images and their effects on the analysis results of the brain activity. ......................................................... 2

2.1 The data flow in the MRI scanner. .......................................................... 8

3.1 A 2D illustration of the susceptibility-induced distortions. The displacement field is along the PE (horizontal) direction and is expressed in terms of the number of voxels shifted. .................. 24

3.2 An example of HySCO results with additional artifacts (red arrows) on the top of the brain. ................................................................. 26

3.3 The block diagram of the coarse-to-fine optimization scheme. The displacement field is estimated at each level of data representation. 29

3.4 Distorted EPI slices overlaying on a T1w image with three different orientation views. The blue lines (cross-hairs) indicate the intersection point of the three views. See the electronic color images. .... 34

3.5 Examples of visual stimuli presented to the subjects during scanning. 35
3.6 An example of a phase-map in the coronal plane. The red line marks the outer boundary of white matter. Note that the color coding represents the position in the visual field (see the color wheel), not the strength of responses, as typical in phase-encoded retinotopic mapping. .................................................. 37

3.7 Blurriness measurements of an original 3D $T_{1w}$ image and two blurred images, which are produced by two Gaussian smoothing filters with different standard deviations $\sigma$. .................................................. 38

3.8 Uncorrected images and their corrected versions created using three SAC methods and corresponding $T_{1w}$ images. Top row: images of a subject in the 3T dataset. Bottom row: images of a subject in the 7T dataset. The arrows point to the artifacts produced by HySCO. .................................................. 41

3.9 Cumulative distributions of the blurriness measurements of the two datasets for five cases: uncorrected, TOPUP, HySCO, TISAC, and Gaussian filtering of the uncorrected data. .......................... 42

3.10 Box-plots of the MI coefficients between structural $T_{1w}$ and fMRI images. .................................................. 43

3.11 Phase-maps projected onto the $T_{1w}$ image of uncorrected and corrected data in the 7T data of Subject 1. Top row: phase-maps in the coronal view. Bottom row: phase-maps in the axial view. The arrows point to the areas with large distortions. See the electronic color images. .................................................. 43

3.12 Phase-maps projected onto the $T_{1w}$ image of uncorrected and corrected data in the 3T data of Subject 4. See the electronic color images. .................................................. 44

3.13 Phase-maps projected onto the $T_{1w}$ image of uncorrected and corrected data in the 7T data of Subject 3. See the electronic color images. .................................................. 44
3.14 Mean percentage of activated voxels in gray matter (GM) and white matter (WM). The error bar shows the corresponding standard deviation of the percentage. 44

4.1 The proposed learning framework (S-Net) for correcting the SAs in reversed-PE images. S-Net accepts a pair of reversed-PE images and produces the displacement field and the corrected images. 52

4.2 The convolutional neural network for mapping a pair of reversed-PE images to the displacement field. Box: output feature maps of a layer. Number inside each box: the number of feature maps in the layer. Number below each box: feature map size relative to the full input image size. 53

4.3 Spatial transformation unit for unwarping distorted image via the displacement field. 54

4.4 Validation loss of S-Nets versus the regularization parameter $\lambda$ and learning rate $\alpha$. The value pairs $(\alpha, \lambda)$ are found by the TPE algorithm. The color indicates the value of the validation loss; the star indicates the pair $(\alpha, \lambda)$ with the lowest validation loss. 61

4.5 Representative visual results of SAC methods from the fMRI-7T dataset. Column 1: input uncorrected images. Columns 2, 3, and 4: output corrected images produced by TOPUP, TISAC, and S-Net, respectively. Rows 1 and 2: left-right and right-left phase-encoding EPI images, respectively. Row 3: absolute difference of LR and RL images. 62

4.6 Representative visual results of SAC methods from the HCP 3T datasets. 63
4.7 Comparisons of the accuracy between the S-Net technique versus other two existing SAC methods on the test sets. Left column: similarity between the corrected EPI images. Right column: similarity between the $T_{1w}$ image and corrected EPI images. Because of differences in the datasets, the plots are drawn in different $y$-axis ranges for clarity. ......................................................... 64

5.1 The proposed learning framework (TS-Net) for correcting the SAs in reversed-PE images. The TS-Net accepts a pair of 3D reversed-PE images and produces the displacement fields and the corrected images. ................................................................. 72

5.2 The convolutional encoder-decoder for mapping a pair of reversed-PE images to the displacement fields. Box: output feature maps of a layer. Number inside each box: the number of feature maps in the layer. Number below each box: feature map size relative to the full input image size. ................................................................. 73

5.3 Validation loss of the models trained with three types of similarity loss (MSE, LCC, and LNCC) versus training time (in second) on the four datasets: (a) fMRI-3T; (b) DWI-3T; (c) fMRI-7T; and (d) DWI-7T. Top row: validation loss in terms of MSE. Middle row: validation loss in terms of LCC. Bottom row: validation loss in terms of LNCC. 79

5.4 Validation loss of four trained models versus training time (in second) on three datasets: (a) fMRI-3T; (b) DWI-3T; and (c) DWI-7T. The four models are: (i) TS-Net without batch normalization and with upsampling layer (UL); (ii) TS-Net with instance normalization and UL; (iii) TS-Net with batch normalization (BN) and transposed convolution; and (iv) TS-Net with BN and UL (proposed method). ................................................................. 80
5.5 Validation loss of three trained models versus training time (in second) on three datasets: (a) fMRI-3T; (b) DWI-3T; and (c) DWI-7T. The three models are: (i) TS-Net with 1D distortion model and \( T_{1w} \) guidance; (ii) TS-Net with 3D distortion model and without \( T_{1w} \) guidance; and (iii) TS-Net with 3D distortion model and \( T_{1w} \) guidance (proposed method). 81

5.6 Validation loss of two trained TS-Net models versus training time (in second) on three datasets: (a) DWI-3T; (b) fMRI-7T; and (c) DWI-7T. The two TS-Net models were trained: (i) from scratch; and (ii) using the pre-trained model of the fMRI-3T dataset. 82

5.7 Samples of three predicted displacement fields (in voxel) of TS-Net from the four test sets. In each subfigure, *left image*: displacement field in the left-right (LR) direction; *middle image*: displacement field in the anterior-posterior (AP) direction; and *right image* displacement field in the superior-interior (SI) direction. The dominant phase-encoding dimension (direction) is shown in red text; the other two other dimensions are shown in white text. 83

5.8 Sample visual results of SAC methods from the four test sets. In each subfigure, *Column 1*: input uncorrected images. *Columns 2, 3, 4, and 5*: output corrected images produced by TOPUP, TISAC, S-Net, and TS-Net, respectively. *Rows 1 and 2*: reversed phase-encoding EPI images. *Row 3*: color bar of the absolutely different maps. *Row 4*: absolute difference between the image pair. *Row 5*: corresponding \( T_{1w} \) image of the reversed-PE images and the estimated displacement fields of the compared SAC methods. For visualization, only the displacement field in the phase-encoding direction of TS-Net is shown. *Row 6*: color bar of the displacement fields, in which the number expresses the number of voxels shifted. 84
5.9 Comparisons of the proposed TS-Net versus other three SAC methods in terms of the LNCC-based accuracy on the test sets. Due to differences in LNCC ranges of the datasets, the plots are drawn in different y-axis ranges for clarity. In each box plot, the top line is the maximum LNCC value excluding the outliers; the bottom line is the minimum LNCC value excluding the outliers; the middle line is the median LNCC value; the solid rectangle is the interquartile range of the LNCC values; and the points are the outliers.

A.1 An illustration of a proton spinning on its axis and protons aligned randomly in the absence of a magnetic field.

A.2 Spins in an external magnetic field.

A.3 An example of the net magnetization.

A.4 States of spins during the excitation.

B.1 An example of MIND difference maps between T1w and other MRI types: (a) $T_{1w}$ image of Subject 1; (b) $T_{1w}$ image of Subject 2; (c) EPI image of Subject 1; (d) MIND map of (a) and (b) with MIND score $D = 0.42$; (e) MIND map of (a) and (c) with MIND score $D = 0.24$. A blue color denotes a small difference, a red color denotes a large difference. See the electronic color image.
List of Tables

2.1 Representative SAC methods (S. mod. = pulse sequence modified). 16

3.1 A summary of the datasets used in the experiments. 33

3.2 Similarity measures and execution times (in second) of TISAC configurations: with and without using the BO technique. Methods using BO do not include time for estimating the hyper-parameters. 40

3.3 Comparison with other SAC methods in terms of execution time (in second). 40

3.4 Comparison with other SAC methods in terms of the blurriness introduced. 42

3.5 Change of the BOLD responses after SACs over scans. 45

4.1 A summary of the datasets used in the experiments. 58

4.2 A summary of the training set and test set for each of the three datasets. 59

4.3 Hyper-parameter values used in training the S-Net models for the three datasets. 61

4.4 Statistical comparisons of the EPI-NMI measures between the proposed S-Net versus the uncorrected data and other two existing SAC methods on the test sets. 65
### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5</td>
<td>Statistical comparisons of the anatomical-NMI measures between the proposed S-Net versus the uncorrected data and other two existing SAC methods on the test sets.</td>
<td>65</td>
</tr>
<tr>
<td>4.6</td>
<td>Elapsed time (in second) of SAC methods for correcting a pair of reversed-PE images with a size of $144 \times 168 \times 111$.</td>
<td>66</td>
</tr>
<tr>
<td>4.7</td>
<td>Elapsed time (in second) of SAC methods for unwarping separately two reversed-PE images with a size of $144 \times 168 \times 111$.</td>
<td>66</td>
</tr>
<tr>
<td>4.8</td>
<td>Training time (in second) of the proposed S-Net model on different datasets.</td>
<td>67</td>
</tr>
<tr>
<td>5.1</td>
<td>A summary of the datasets used in the experiments.</td>
<td>77</td>
</tr>
<tr>
<td>5.2</td>
<td>Values of hyper-parameters in training TS-Net on the four datasets.</td>
<td>78</td>
</tr>
<tr>
<td>5.3</td>
<td>A summary of the training set and test set for each of the four datasets.</td>
<td>78</td>
</tr>
<tr>
<td>5.4</td>
<td>Statistical comparisons on the test sets of the LNCC measures between the TS-Net outputs versus four image types: uncorrected images, TOPUP outputs, TISAC outputs, and S-Net outputs.</td>
<td>85</td>
</tr>
<tr>
<td>5.5</td>
<td>Processing time (in second) of SAC methods for correcting a pair of reversed-PE images.</td>
<td>87</td>
</tr>
<tr>
<td>5.6</td>
<td>Training time (in second) of the proposed TS-Net model on different datasets.</td>
<td>87</td>
</tr>
<tr>
<td>A.1</td>
<td>Magnetic properties of nuclei.</td>
<td>95</td>
</tr>
</tbody>
</table>
Acronyms

2-D Two dimensional
3-D Three dimensional
ANOVA One-way analysis of variance
BO Bayssian optimization
BN Batch-normalization
BOLD Blood oxygenation-level dependent
CDF Cumulative distribution function
CNN Convolutional neural network
CSF Cerebrospinal fluid
DWI Diffusion-weighted MRI
EPI Echo planar imaging
fMRI Functional magnetic resonance imaging
FOV Field of view
FSL FMRIB software library
GM Gray matter
<table>
<thead>
<tr>
<th>Acronyms</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP</td>
<td>Gaussian process</td>
</tr>
<tr>
<td>GPR</td>
<td>Gaussian process regressor</td>
</tr>
<tr>
<td>GRE</td>
<td>Gradient echo</td>
</tr>
<tr>
<td>HCP</td>
<td>Human Connectome Project</td>
</tr>
<tr>
<td>HySCO</td>
<td>Hyper-elastic susceptibility artifact correction</td>
</tr>
<tr>
<td>LCC</td>
<td>Local cross-correlation</td>
</tr>
<tr>
<td>LeakyReLU</td>
<td>Leaky rectified linear unit</td>
</tr>
<tr>
<td>LNCC</td>
<td>Local normalized cross-correlation</td>
</tr>
<tr>
<td>LR</td>
<td>Left-to-right</td>
</tr>
<tr>
<td>MHz</td>
<td>Megahertz</td>
</tr>
<tr>
<td>MI</td>
<td>Mutual information</td>
</tr>
<tr>
<td>MIND</td>
<td>Multi-dimensional image descriptor</td>
</tr>
<tr>
<td>MOCO</td>
<td>Motion correction</td>
</tr>
<tr>
<td>MR</td>
<td>Magnetic resonance</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>ms</td>
<td>Millisecond</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean squared error</td>
</tr>
<tr>
<td>NCC</td>
<td>Normalized cross-correlation</td>
</tr>
<tr>
<td>NCC</td>
<td>Normalized gradient field</td>
</tr>
<tr>
<td>NMI</td>
<td>Normalized mutual information</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance</td>
</tr>
</tbody>
</table>
## Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAV</td>
<td>Percentage of activated voxels</td>
</tr>
<tr>
<td>PE</td>
<td>Phase-encoding</td>
</tr>
<tr>
<td>PSF</td>
<td>Point spread function</td>
</tr>
<tr>
<td>RF</td>
<td>Radio-frequency</td>
</tr>
<tr>
<td>RL</td>
<td>Right-to-left</td>
</tr>
<tr>
<td>SAs</td>
<td>Susceptibility artifacts</td>
</tr>
<tr>
<td>SAC</td>
<td>Susceptibility artifact correction</td>
</tr>
<tr>
<td>SPM</td>
<td>Statistical parametric mapping</td>
</tr>
<tr>
<td>SSD</td>
<td>Sum of squared difference</td>
</tr>
<tr>
<td>STU</td>
<td>Spatial transform unit</td>
</tr>
<tr>
<td>T</td>
<td>Tesla</td>
</tr>
<tr>
<td>$T_{1w}$</td>
<td>T1 weighted</td>
</tr>
<tr>
<td>$T_{1wR}$</td>
<td>$T_{1w}$-based regularization</td>
</tr>
<tr>
<td>$T_{2w}$</td>
<td>T2 weighted</td>
</tr>
<tr>
<td>TE</td>
<td>Echo time</td>
</tr>
<tr>
<td>TISAC</td>
<td>Inverse-gradient susceptibility artifact correction using $T_{1w}$ image</td>
</tr>
<tr>
<td>TPE</td>
<td>Tree of Parzen estimator</td>
</tr>
<tr>
<td>TR</td>
<td>Time repetition</td>
</tr>
<tr>
<td>WM</td>
<td>White matter</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Chapter contents

| 1.1 | Research motivation and objectives | 1 |
| 1.2 | Thesis organization | 3 |
| 1.3 | Research contributions | 4 |
| 1.4 | Publications | 6 |

1.1 Research motivation and objectives

Echo planar imaging (EPI) is a fast magnetic resonance imaging technique that provides images of living species in only a fraction of a second [1, 2]. The fast imaging capability enables EPI to record rapid changes of physiological processes in the human body, e.g., brain activity and heartbeat. Thus, EPI is a widely used technique for human brain studies in both clinical diagnosis and scientific investigation. For example, EPI is used to produce diffusion-weighted imaging (DWI) images for evaluation of stroke, and functional MRI (fMRI) images for localization of brain activity [1].

Despite its speed, EPI is prone to distortions due to local field inhomogeneities, which are caused by the differences in magnetic susceptibility of various imaged tissues (e.g., fat versus blood) [3, 4]. These local field inhomogeneities affect the spatial encoding of the signal. Consequently, they degrade the acquired images
1.1. Research motivation and objectives

by geometrical deformations (stretching and compressing) and intensity modulations [5]. These distortions are known as susceptibility artifacts (SAs). The SAs are more severe at high field strengths [6, 7] and in rapid imaging techniques such as EPI [3, 8]. These artifacts can be easily seen in the interface regions, particularly between the cerebral cortex and non-brain areas [4]. In EPI images, SAs are most noticeable along the phase-encoding (PE) direction. They constantly appear reversed in two reversed-PE images which are acquired using identical EPI sequences but with inverse PE directions\(^a\) [9, 10, 11]. Figs. 1.1(a) and 1.1(b) illustrate an example of the susceptibility artifacts in slices from two functional MRI images acquired using the identical EPI sequences but with opposite PE directions from left-to-right (LR) and right-to-left (RL). In Fig. 1.1(a), some areas (red arrows) are compressed, while other areas (cyan arrows) are stretched. The distortion patterns in Fig. 1.1(a) are reversed with the distortion patterns in Fig. 1.1(b).

\(^a\)In MRI, the phase encoding direction is also known as the polarity of the PE gradient or the blip.

Figure 1.1: Example of the susceptibility artifacts in a pair of LR and RL phase encoding EPI images and their effects on the analysis results of the brain activity.

In brain activity studies, the analysis of the functional images is always overlaid with a structural MRI image as it has a better spatial resolution [12]. The SAs disrupt the geometric correspondence between functional and structural data. This disruption subsequently leads to misplacements of detected activation patterns in fMRI studies. Fig. 1.1(c) shows a map indicating the cortical activation pattern (see color areas), which is achieved by analyzing an uncorrected LR fMRI
dataset. The thin blue lines mark the boundary between white matter and gray matter. In the example, there is a significant amount of activity located incorrectly in the white matter because of geometric distortions in fMRI images.

Correcting the susceptibility artifacts is an important step in the preprocessing pipeline of brain studies using EPI [7, 13]. Susceptibility artifact correction (SAC) is a very challenging task for two main reasons. First, there is no correction ground-truth of the acquired EPI images. Second, performing SAC on brain EPI images requires a high degree of correction accuracy since the brain structure is relatively small and complex. Many SAC methods have been developed to address these challenges. However, the existing SAC methods require high computational resources, modified scanner hardware, or a modified acquisition protocol [14].

The overall goal of this project is to develop algorithms for susceptibility artifact correction in brain EPI images. Our approach allows correcting the susceptibility artifacts without changing the MRI scanner and the acquisition protocol. The novelty of this project is the ability to correct the susceptibility artifacts with high correction accuracy within a short time. This project is a step towards correcting the SAs in real-time and integrating the SAC into MRI scanners. More specifically, the aims of this project can be broken down into:

- Investigate and evaluate approaches to correct the susceptibility artifacts in EPI images, especially the approach of using a pair of reversed-PE images.
- Develop algorithms that correct the susceptibility artifacts in brain EPI images with a high correction accuracy and at a high speed.

1.2 Thesis organization

This thesis consists of six chapters:

- Chapter 1 outlines the project motivations and objectives. It highlights the research contributions and publications.
1.3. Research contributions

- Chapter 2 presents the fundamentals of the susceptibility artifacts in MRI. It also provides a systematic overview of existing susceptibility artifact correction methods, in five major categories: fieldmap based, point spread function based, reversed phase-encoding based, image-registration based, and hybrid methods. The strengths and weaknesses of each category are discussed.

- Chapter 3 presents a new iterative-optimization SAC algorithm using a $T_1$ weighted regularization. The proposed algorithm uses the structural information of the $T_{1w}$ image to regularize the corrected images and select the regularization parameters automatically. This chapter also introduces novel techniques to assess the validity of the corrected images.

- Chapter 4 presents the first deep learning method to correct the susceptibility artifacts in EPI images. The proposed method uses the unsupervised manner to train the network in an end-to-end setting.

- Chapter 5 presents a new multi-modal deep learning method to correct the susceptibility artifacts in EPI images in an end-to-end setting. The proposed network is designed to cope with different image sizes, resolutions, and modalities. In training the network, $T_1$ and EPI modalities are used.

- Chapter 6 summarizes the research activities and provides the concluding remarks.

1.3 Research contributions

The principal contributions of this thesis are in the fields of biomedical imaging, machine learning, and image processing. The contributions are listed as follows:

- A review of the state-of-the-art SAC methods is provided. The survey SAC methods include five categories: fieldmap, point spread function, reversed phase-encoding, image-registration, and hybrid methods.
1.3. Research contributions

- A novel iterative-optimization algorithm using $T_{1w}$ regularization is proposed for correcting the susceptibility artifacts a pair of reversed-PE images. The structural information of the $T_{1w}$ image is used to control the reality of corrected images and select the regularization parameters automatically. The proposed algorithm produces corrected images by solving an optimization problem using the Gauss-Newton method. The regularization parameters are selected automatically using Bayesian optimization. The proposed algorithm is validated on two functional MRI datasets and also via the results of the post-processing fMRI step.

- A novel unsupervised deep learning method is proposed for SAC in an end-to-end setting. The proposed network consists of two components: (i) a convolutional encoder-decoder to map a reversed-PE image pair to the displacement field; and (ii) a spatial transform unit to unwar the input images and produce the corrected images. The proposed network is trained in an unsupervised manner, without ground-truth data. For a new pair of reversed-PE images, the displacement field and corrected images are obtained simultaneously by evaluating the trained model. Thus, the proposed method can avoid the iterative optimization, and thereby reducing the processing time significantly.

- A novel multi-modal deep learning method is proposed for correcting the susceptibility artifacts in a reversed-PE image pair. The proposed network also consists of two components: (i) a convolutional encoder-decoder to map the reversed-PE image pair to the displacement field; and (ii) a spatial transform unit to unwar the input images and produce the corrected images. The encoder-decoder is designed so that the trained model can cope with different input image sizes; in the training phase, EPI and $T_1$ modalities are used. In the inference phase, only the input EPI images are fed into the trained model.
1.4 Publications

The publications arising from this project (June 2016 - April 2020) are listed as follows:


Background and Related Work

Chapter contents

2.1 Fundamentals of susceptibility artifacts in MRI . . . . . . . . . . 8
  2.1.1 Susceptibility artifacts in conventional 2D GRE imaging . 10
  2.1.2 Susceptibility artifacts in EPI . . . . . . . . . . . . . . . . . 13

2.2 Existing susceptibility artifact correction methods . . . . . . . 15
  2.2.1 Fieldmap-based SACs . . . . . . . . . . . . . . . . . . . . . 15
  2.2.2 Point spread function based SACs . . . . . . . . . . . . . . 17
  2.2.3 Reversed-PE based SACs . . . . . . . . . . . . . . . . . . . 17
  2.2.4 Image registration based SACs . . . . . . . . . . . . . . . . 18
  2.2.5 Hybrid SACs . . . . . . . . . . . . . . . . . . . . . . . . . . 19

2.3 Chapter summary . . . . . . . . . . . . . . . . . . . . . . . . . . . 20

The key contribution of this thesis is the development of techniques that are capable of correcting the distortions, i.e. susceptibility artifacts, in the brain EPI images. To fulfil this goal, we have to understand the fundamentals of the susceptibility artifacts, their effects, and the existing susceptibility artifact correction methods. This chapter discusses the necessary background materials and related works on this line of investigation.
2.1 Fundamentals of susceptibility artifacts in MRI

Magnetic resonance imaging is an invasive technique for producing images of the inside of the biological samples. All MRI techniques rely on a physical phenomenon called nuclear magnetic resonance (NMR) to generate the MR signals and subsequently generate the image of the object being scanned [2, 15, 16]. The magnetic resonance (MR) image is not a photograph of the sample. It is a map of some properties related to the sample, such as the net magnetization or the density of spins. Appendix A describes in more detail about MR signal generation and image formation.

In general, MRI generates MR signals by applying a pulse sequence, which is the series of oscillating radio-frequency (RF) pulses and changing magnetic field gradients. The emitted MR signals are then sampled and saved in special storage known as the \textit{k-space}. The MR image is formed by applying the inverse Fourier to the \textit{k-space} data. Fig. 2.1 demonstrates the flow of data in the MR image formation.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{mr_image_flow.png}
\caption{The data flow in the MRI scanner.}
\end{figure}

In MRI, 2D imaging is the technique of choice for most applications. The 2D imaging technique constructs a 3D image from a set of 2D slices. The data acquired in the \textit{k-space} is Fourier transform data of a 2D slice from the scanned object. As shown in [16], in the homogeneity field, the MR signal at time \( t \) is written as

\[
S(t) = \int \int M(x, y) e^{-j\gamma \int_0^t (G_x(\tau)x + G_y(\tau)y) d\tau} dx dy, \tag{2.1}
\]

where \( M(x, y) \) denotes the total net magnetization of the volume at location \((x, y)\).
2.1. Fundamentals of susceptibility artifacts in MRI

$G_x(\tau)$ is the time-varying field of the frequency-encoding (readout) gradient along the $x$ direction at the given instant $\tau$. $G_y(\tau)$ is the time-varying field of the phase-encoding gradient along the $y$ direction.

The MR signal is sampled and filled in the 2D k-space. Let’s set the changes in k-space over time as

$$
\begin{align*}
    k_x(t) &= \frac{\gamma}{2\pi} \int_0^t G_x(\tau) d\tau, \\
    k_y(t) &= \frac{\gamma}{2\pi} \int_0^t G_y(\tau) d\tau.
\end{align*}
$$

The recorded data at a given point $(k_x, k_y)$ in the k-space can be written as

$$
S(k_x, k_y) = \int \int M(x, y) e^{-i2\pi(k_xx + k_yy)} dx dy.
$$

Eq. (2.4) indicates that the acquired data in the k-space is the Fourier transform of the net magnetization map. Therefore, the reconstructed image, which reveals the map of the net magnetization, is the inverse Fourier transform of data in the k-space.

It is worth noting that under the ideal field homogeneity, the reconstructed image is exactly the inverse Fourier transform of data in the k-space. However, in reality, it is hard to maintain the field homogeneity in the scanning environment. There are many sources causing field inhomogeneities in the static magnetic field. The main and unavoidable source of the field inhomogeneities comes from magnetic susceptibility differences of local imaged tissues.

Magnetic susceptibility of a tissue is a measure that reveals the magnetized capability of the tissue in a strong magnetic field. Different tissue types have different magnetic susceptibilities. Thus, in the volume of space containing tissues of the closed types, the local field generated by the differences in susceptibility of tissues is small or zero. However, in the volume of space containing the interface, e.g. air versus bone and fat versus blood, the differences in susceptibility are significant enough to set up a local magnetic field [3, 4]. The variety of local
magnetic fields over all the scanned space leads to the field inhomogeneities.

Let $B(x, y)$ denote the field inhomogeneity at location $(x, y)$. In the presence of the field inhomogeneities, the MR signal in Eq. (2.1) becomes

$$S(t) = \int \int M(x, y) e^{-i \gamma \int_0^t (G_x(\tau)x + G_y(\tau)y) d\tau} e^{-i \gamma B(x, y)t} dx\, dy.$$  \hspace{1cm} (2.5)

Thus, in the presence of the field inhomogeneities, the reconstructed image by applying the inverse Fourier transform no longer reflects exactly the map of the net magnetization. Instead, the reconstructed image reflects the net magnetization with some artifacts, i.e. geometric distortions and intensity modulations [5, 16]. Those distortions are known as susceptibility artifacts. Sections 2.1.1 and 2.1.2 mathematically explain the effects of the field inhomogeneities in the conventional gradient-echo (GRE) imaging and echo planar imaging, respectively.

### 2.1.1 Susceptibility artifacts in conventional 2D GRE imaging

Gradient-echo pulse sequence is a common MRI technique as it reduces the imaging time and decreases motion artifacts [17]. As a conventional 2D MRI technique, 2D GRE imaging generates data for a slice using multiple excitation pulses. Each excitation pulse helps to generate data for a $k_x$ line in the k-space. In the conventional 2D GRE imaging, the acquisition time $t$ in Eq. (2.5) can be written as

$$t = t_{TE} + \frac{2\pi k_x}{\gamma G_x},$$  \hspace{1cm} (2.6)

where $t_{TE}$ is the echo time, which is the constant interval between applying the RF pulse and starting the signal acquisition. $G_x$ is the amplitude of the readout gradient.

Thus, the recorded data at any point $(k_x, k_y)$ in the k-space in the presence of the field inhomogeneities can be written as

$$S(k_x, k_y) = \int \int M(x, y) e^{-i 2\pi (k_x x + k_y y)} e^{-i \gamma B(x, y)(t_{TE} + \frac{2\pi k_y}{\gamma G_x})} dx\, dy,$$  \hspace{1cm} (2.7)
or

\[
S(k_x, k_y) = \int \int M(x, y) e^{-iyB(x, y)T_E} e^{-i2\pi(k_x(x + \frac{B(x, y)}{G_x}) + k_y y) dx dy.} \tag{2.8}
\]

Eq. (2.8) indicates that in the conventional GRE imaging, the field inhomogeneities effects are along the frequency-encoding (x) direction.

To find the inverse Fourier of the signal \( S \), we define a coordinate transformation operator as

\[
\tilde{x} = T(x) = x + \frac{B(x, y)}{G_x}, \tag{2.9}
\]

where \( T(x) \) is the operator that transforms the undistorted coordinate \( x \) to the distorted coordinate \( \tilde{x} \). If the local field inhomogeneity \( B(x, y) \) is smaller than the phase-encoding gradient strength \( G_x \), \( T(x) \) increases or decreases monotonically in the \( x \) direction; thus, \( T \) is invertible. We can obtain:

\[
x = T^{-1}(\tilde{x}), \tag{2.10}
\]

\[
d\tilde{x} = 
\left(1 + \frac{\partial}{\partial x} \frac{B(x, y)}{G_x} \right) dx = 
\left(1 + \frac{\partial}{\partial T^{-1}(\tilde{x})} \frac{B(T^{-1}(\tilde{x}), y)}{G_x} \right) dx = P(T^{-1}(\tilde{x}), y) dx. \tag{2.11}
\]

Substituting \( \tilde{x} \) for \( x \) in Eq. (2.8) using Eqs. (2.10) and (2.11), we obtain:

\[
S(k_x, k_y) = \int \int \frac{M(T^{-1}(\tilde{x}), y)}{P(T^{-1}(\tilde{x}), y)} e^{-iyB(T^{-1}(\tilde{x}), y)T_E} e^{-i(k_x \tilde{x} + k_y y)} d\tilde{x} dy. \tag{2.12}
\]

After 2D inverse Fourier transform of \( S(k_x, k_y) \), the measured GRE image density \( I_{GRE}(u, v) \) is obtained:

\[
I_{GRE}(u, v) = \int \int S(k_x, k_y) e^{i2\pi(k_x u + k_y v)} dk_x dk_y
= \int \int \int \int \frac{M(T^{-1}(\tilde{x}), y)}{P(T^{-1}(\tilde{x}), y)} e^{-iyB(T^{-1}(\tilde{x}), y)T_E}
\cdot e^{-i2\pi(k_x \tilde{x} + k_y y)} e^{i2\pi(k_x u + k_y v)} dk_x dk_y d\tilde{x} dy
= \int \int \left\{ \frac{M(T^{-1}(\tilde{x}), y)}{P(T^{-1}(\tilde{x}), y)} e^{-iyB(T^{-1}(\tilde{x}), y)T_E} \right\}
\cdot \left\{ \int e^{-i2\pi k_x (\tilde{x} - u)} dk_x \right\} \left\{ \int e^{-i2\pi k_y (y - v)} dk_y \right\} d\tilde{x} dy. \tag{2.13}
\]
We have
\[ \int e^{-i2\pi f t} \, df = \delta(t) \quad \text{where} \quad \delta(t) = \begin{cases} 1, & t = 0, \\ 0, & t \neq 0. \end{cases} \quad (2.14) \]

Thus,
\[ \int e^{-i2\pi k_x(x-u)} \, dk_x = \delta(x-u) \quad \text{and} \quad \int e^{-i2\pi k_y(y-v)} \, dk_y = \delta(y-v). \quad (2.15) \]

The reconstructed image is now written as
\[ I_{\text{GRE}}(u, v) = \int \int \left\{ \frac{M(T^{-1}(\tilde{x}), y)}{P(T^{-1}(\tilde{x}), y)} \right\} e^{-i\gamma B(T^{-1}(\tilde{x}), y)t_{\text{TE}}} \delta(x-u) \delta(y-v) \, d\tilde{x} \, dy \]
\[ = \int \int \frac{M(T^{-1}(u), v)}{P(T^{-1}(u), v)} e^{-i\gamma B(u,v)t_{\text{TE}}} \, d\tilde{x} \, dy \]
\[ = \frac{M(T^{-1}(u), v)}{P(T^{-1}(u), v)} e^{-i\gamma B(u,v)t_{\text{TE}}} \]
\[ = \left(1 + \frac{\partial}{\partial T^{-1}(u)} B(T^{-1}(u), v) \right)^{-1} M(T^{-1}(u), v) e^{-i\gamma B(u,v)t_{\text{TE}}}, \quad (2.16) \]

or
\[ I_{\text{GRE}}(\tilde{x}, y) = \left(1 + \frac{\partial}{\partial \tilde{x}} B(x, y) \right)^{-1} M(x, y) e^{-i\gamma B(x,y)t_{\text{TE}}}. \quad (2.17) \]

Eq. (2.17) demonstrates that in the presence of the field inhomogeneities, the reconstructed MRI image is distorted. The artifacts include the intensity modulations which are represented by \( \left(1 + \frac{\partial}{\partial \tilde{x}} B(x, y) \right)^{-1} \). The artifacts also include the geometric distortions which are represented by the shifting of net magnetization in the \( x \) (readout) direction. In other words, the intensity of the corrected image at \((x, y)\) is shifted to \((\tilde{x}, y)\) in the acquired image.

Eq. (2.17) also shows that the strength of distortions is proportional to the magnitude of the local field inhomogeneities, and it is inversely proportional to the amplitude of the readout gradient. In general, field inhomogeneities are proportional to the field strength. Therefore, in the conventional GRE imaging, the distortions are increased at higher field strengths and lower amplitudes of the readout gradient.
We summarize the properties of susceptibility artifacts in conventional GRE images as follows:

- The artifacts include both intensity modulations and geometric distortions. In practice, the susceptibility artifacts in conventional GRE images can be negligible.
- The artifacts are severe in high field strengths and high spatial resolutions.
- The geometric distortions occur mainly in the **readout** direction. They appear inversely in the image acquired with the reversed readout gradient direction as the sign of the local field inhomogeneities depends on the sign of the readout gradient.

### 2.1.2 Susceptibility artifacts in EPI

Echo planar imaging invented by Peter Mansfield in 1977 is an ultrafast MRI technique [18]. The key idea of EPI is that the data of an entire slice are acquired by a single excitation pulse instead of multiple pulses as in conventional imaging methods. Thus, a volume can be acquired within a few seconds (e.g. 2s for a volume size of $128 \times 128 \times 132$ with a 3T scanner).

In EPI, as shown in [19, 20], the acquisition time $t$ in Eq. (2.5) can be written as

$$t = t_{TE} + \frac{2\pi k_x}{\gamma G_x} + \frac{2\pi k_y T_y}{\gamma G_y \tau_y}, \quad (2.18)$$

where $T_y$ is the time interval between adjacent $k_y$-lines in the k-space, $\tau_y$ is the duration of applying the increment phase-encoding gradient, and $G_y$ is the average phase-encoding gradient field of duration $\tau_y$. Therefore, the measured data at any point $(k_x, k_y)$ in the k-space in the presence of the local field inhomogeneities $B(x, y)$ can be written as

$$S(k_x, k_y) = \int \int M(x, y) e^{-i2\pi(k_x x + k_y y)} e^{-\gamma B(x, y)(t_{TE} + \frac{2\pi k_x}{\gamma G_x} + \frac{2\pi k_y T_y}{\gamma G_y \tau_y})} dx dy, \quad (2.19)$$
2.1. Fundamentals of susceptibility artifacts in MRI

or

\[ S(k_x, k_y) = \int \int M(x, y) e^{-i\gamma B(x,y)t_{TE}} e^{-i2\pi(k_x(x + B(x,y)t_{TE}) + k_y(y + B(x,y)\tau_y))} \, dx \, dy. \]  \hspace{1cm} (2.20)

The EPI technique often uses a very high amplitude of gradient \( G_x \) \cite{21}. Thus, the distortion term \( \frac{B(x,y)}{\gamma} \) in the readout direction is relatively small. In the following, we assume that there is no geometric distortion in the readout direction. Eq. (2.20) can be simplified as

\[ S(k_x, k_y) = \int \int M(x, y) e^{-i\gamma B(x,y)t_{TE}} e^{-i2\pi(k_x x + k_y(y + \frac{B(x,y)\tau_y}{\gamma}))} \, dx \, dy. \]  \hspace{1cm} (2.21)

In a similar derivation as in Eqs. (2.9) to (2.11), we can define

\[ \bar{y} = R(y) = y + \frac{B(x,y)\tau_y}{G_y \tau_y}, \]  \hspace{1cm} (2.22)

\[ y = R^{-1}(\bar{y}), \]  \hspace{1cm} (2.23)

\[ d\bar{y} = \left(1 + T_y \frac{\partial}{\partial y} B(x,y) \right) \left(1 + T_y \frac{\partial}{\partial \tau_y} B(x,R^{-1}(\bar{y})) \right) d\bar{y}. \]  \hspace{1cm} (2.24)

Substituting \( \bar{x} \) for \( x \) in Eq. (2.21) using Eqs. (2.23) and (2.24), we obtain:

\[ S(k_x, k_y) = \int \int M(x, R^{-1}(\bar{y})) e^{-i\gamma B(x,R^{-1}(\bar{y}))t_{TE}} e^{-i2\pi(k_x x + k_y \bar{y})} \, dx \, d\bar{y}. \]  \hspace{1cm} (2.25)

By applying an inverse Fourier transform to \( S(k_x, k_y) \) similarly as in the conventional GRE imaging, we obtain the measured EPI image as

\[ I_{\text{EPI}}(x, \bar{y}) = \left(1 + T_y \frac{\partial}{\partial y} B(x,y) \right)^{-1} M(x, y) e^{-i\gamma B(x,y)t_{TE}}. \]  \hspace{1cm} (2.26)

In contrast to the GRE imaging, Eq. (2.26) shows that the field inhomogeneities in EPI mainly affect the lines along \( k_y \) in the k-space. The measured EPI image becomes severely distorted in the phase-encoding direction. Since \( \tau_y \ll T_y \) and the amplitude of the phase-encoding gradient increment is usually small, the distortions in EPI are not negligible.
2.2. Existing susceptibility artifact correction methods

We conclude the properties of susceptibility artifacts in EPI images as follows:

- The artifacts include both intensity modulations and geometric distortions. In practice, the susceptibility artifacts in EPI images are not negligible.
- The artifacts are severe in high field strengths and high spatial resolutions.
- The geometric distortions occur mainly in the phase-encoding direction. They appear inversely in the image acquired with the reversed phase-encoding direction as the sign of the local field inhomogeneities depends on the sign of the phase-encoding.

2.2 Existing susceptibility artifact correction methods

Several SAC methods have been proposed for multiple types of MRI, such as structural MRI, DWI, and fMRI. In general, they can be divided into five categories: (i) fieldmap-based; (ii) point spread function (PSF) based; (iii) image registration based; and (iv) inverse phase-encoding based methods. Table 2.1 summarizes the SAC methods discussed below.

2.2.1 Fieldmap-based SACs

Fieldmap-based SAC methods estimate phase dispersions caused by the field inhomogeneities. The estimated phase dispersion over the entire scanned view is called the fieldmap. An early approach derives the fieldmap from two complex MRI images acquired by different values of echo time (TE) [10]. Another approach requires modified MRI sequences to produce the fieldmap quickly [19, 24, 25].

After the fieldmap is estimated, the corrected images can be obtained by unwarping the distorted images [9, 22], or rewinding the additional accumulated phase in the k-space [23], thereby obtaining the corrected images. There have been multiple approaches to estimate the fieldmap. However, the main limitation of unwarping in the image space is the lack of intensity correction. Rewinding
2.2. Existing susceptibility artifact correction methods

Table 2.1: Representative SAC methods (S. mod. = pulse sequence modified).

<table>
<thead>
<tr>
<th>Category</th>
<th>S. mod.</th>
<th>Authors</th>
<th>Year</th>
<th>Datatype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fieldmap</td>
<td>No</td>
<td>Jezzard and Balaban</td>
<td>1995</td>
<td>EPI</td>
<td>Use the fieldmap derived from complex images acquired by different TEs to unwarp the distorted images.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reber et al.</td>
<td>1998</td>
<td>EPI</td>
<td>Smooth the displacement derived by the method in [27] using a 2D Gaussian kernel to increase the signal-to-noise ratio.</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Hutton et al.</td>
<td>2002</td>
<td>fMRI</td>
<td>Derive the fieldmap from EPI images acquired with different TEs.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kadah and Hu</td>
<td>1997</td>
<td>EPI</td>
<td>Use the fieldmap to rewind the additional accumulated phase in the k-space (called SPHERE).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wan et al.</td>
<td>1997</td>
<td>EPI</td>
<td>Calculate the fieldmap using a set of reference scans generated by turning off the PE gradient of the EPI pulse sequence.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chen and Wyrwicz</td>
<td>1999</td>
<td>EPI</td>
<td>Incorporate a set of fieldmaps by the multi-channel modulation algorithm to obtain corrected images.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cachayipoo et al.</td>
<td>2008</td>
<td>EPI</td>
<td>Derive the fieldmap from EPI images with modified k-space trajectories.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Robson et al.</td>
<td>1997</td>
<td>EPI</td>
<td>Measure the PSF by an EPI sequence with added PE gradients, constant time but variable magnitude.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Munger et al.</td>
<td>2000</td>
<td>EPI</td>
<td>Unwarp the distorted image given the measured PSF by a conjugate gradient algorithm.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zeng and Constable</td>
<td>2002</td>
<td>EPI</td>
<td>Correct both the intensity and geometric distortions in EPI images by measured PSF as in [26].</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Zaitsev et al.</td>
<td>2004</td>
<td>EPI</td>
<td>Measure the PSF by integrating a parallel imaging technique into the acquisition.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In and Speck</td>
<td>2012</td>
<td>EPI</td>
<td>Propose a method to obtain a kernel for distortion correction from the full PSF.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kybic et al.</td>
<td>2000</td>
<td>EPI</td>
<td>Register distorted EPI images by modelling the displacement with splines and using the SSD similarity measure.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Studholme et al.</td>
<td>2000</td>
<td>fMRI</td>
<td>Register EPI images using a multimodality non-rigid registration algorithm with log-intensity measure.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wu et al.</td>
<td>2006</td>
<td>fMRI</td>
<td>Register distorted images based on Thirion’s demons.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wu et al.</td>
<td>2008</td>
<td>EPI</td>
<td>Register distorted EPI images to a T$^2_w$ using mutual information.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Chang and Fitzpatrick</td>
<td>1992</td>
<td>Structural</td>
<td>Introduce the theoretical justification of the correction using inverse phase-encoded images; correct each 1D image along the PE direction independently by finding pairs of corresponding points in the given two images.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Andersson et al.</td>
<td>2003</td>
<td>DWI</td>
<td>Model the displacement as a function of discrete cosine basis functions (called TOPUP).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Holland et al.</td>
<td>2010</td>
<td>fMRI</td>
<td>Model the reversed-PE SAC as a diffusion registration problem.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ruthotto et al.</td>
<td>2012</td>
<td>DWI</td>
<td>Introduce an additional non-linear regularizer into the diffusion regularized problem (called HySCO).</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Irfanoglu et al.</td>
<td>2015</td>
<td>DWI</td>
<td>Incorporate a T$^2_w$ image into the reversed-PE registration.</td>
</tr>
<tr>
<td>Reversed-PE</td>
<td>No</td>
<td>Xiang and Ye</td>
<td>2007</td>
<td>EPI</td>
<td>Modify the PSF method to generate the fieldmap in short time.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gholipour et al.</td>
<td>2007</td>
<td>EPI</td>
<td>Introduce a correction method that uses the algorithm in [23] to rewind the phase difference generated by Xiang and Ye.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ghofipour et al.</td>
<td>2011</td>
<td>EPI</td>
<td>Use the fieldmap as a guide for registering the distorted EPI images to the undistorted T$^2_w$ structural MRI.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daga et al.</td>
<td>2014</td>
<td>EPI</td>
<td>Propose a phase unwrapping algorithm that produces the initialization to register the distorted EPI image to the T$^1_w$ image.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gelman et al.</td>
<td>2014</td>
<td>EPI</td>
<td>Use the fieldmap and reversed-PE approaches to correct the geometric distortions and intensity distortions, respectively.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bao et al.</td>
<td>2016</td>
<td>EPI</td>
<td>Use the reversed-PE approach to eliminate the distortions caused by the field inhomogeneities in the fieldmap.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In et al.</td>
<td>2015</td>
<td>EPI</td>
<td>Correct the two reversed-PE images by PSF; introduce the corrected image as the weighted mean image of the two corrected PSF images.</td>
</tr>
</tbody>
</table>

phase in the k-space allows both geometric and intensity corrections but typically requires customized sequences.
2.2. Existing susceptibility artifact correction methods

2.2.2 Point spread function based SACs

PSF-based SAC methods consider an acquired image as a convolution between the “true” image with a PSF. By estimating the PSF of the system, the undistorted image can be reconstructed. A PSF estimation technique based on constant time imaging was first introduced by Robson et al. for correcting EPI distortions and quantifying the MRI degradation [26]. Subsequently, the PSF estimation was adopted to correct EPI distortions [20, 27].

Zaitsev et al. proposed a further optimized PSF estimation by integrating parallel imaging into the acquisition to correct distortions faster and more reliably, even at high field strengths [28]. In and Speck proposed a method to obtain a kernel for distortion correction from the full PSF to improve the correction accuracy of the SAC [29]. PSF-based SAC methods can correct both geometric distortions and intensity modulations; however, they require the MRI scanner to support configurable MRI sequences.

2.2.3 Reversed-PE based SACs

Reversed-PE based SAC methods utilize two reversed-PE images to estimate the displacement field over the image domain. The corrected images are obtained by unwarping the distorted images by the estimated displacement field. Chang and Fitzpatrick initially introduced the theoretical justification of correcting the SAs using reversed-PE structural images [5]. They then proposed a “cumulative line-integral” method to find the corresponding points, which are used to determine the displacement in two corresponding lines along the PE direction of the given reversed-PE images. Bowtell et al. implemented the original reversed-PE method for 2D EPI [44].

The corrections of the method proposed in [5] are not smooth since the method estimates the displacement in each line along the PE direction independently, without considering surrounding lines. To estimate the displacement field, An-
dersson et al. proposed an alternative approach that considers the displacement at a pixel as a function of discrete cosine basis functions to construct an objective function [34]. This method is called TOPUP and is integrated into the FSL package\(^a\).

Several methods have considered the reversed-PE based SAC as a registration problem. The two distorted reversed-PE images are registered so that their corrected versions are as similar as possible. Holland et al. integrated the reversed-PE approach into a registration framework to correct SAs [11]. Ruthotto et al. combined the registration framework and a constraint inspired by the hyper-elastic image registration to achieve more realistic corrections [35, 45]. This method is called HySCO, and its implementation is included in the SPM12 toolbox\(^b\). Another approach combines an independent image, specifically a T\(_2\) weighted image (T\(_2w\)) image, into the reversed-PE registration to regularize corrections [36].

Reversed-PE based SAC methods can correct both geometric and intensity distortion. They outperform fieldmap and image registration based methods in terms of geometrical correction fidelity, as shown in [46]. The reversed-PE based approach is the most common SAC method, and it is used to correct the fMRI data in the biggest MRI neuroimaging dataset - the Human Connectome Project (HCP) [47]. However, compared to other SAC approaches, registering reversed-PE images requires many constraints, such as the smoothness of the displacement field and the alignment of the correction to the structural image. The reversed-PE methods may produce less meaningful and blurred corrections if unsuitable constraints are used.

### 2.2.4 Image registration based SACs

Image registration based SAC methods map the distorted EPI images to a reference image using a non-rigid model. These methods usually estimate displacements

---

\(^a\)https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/topup

\(^b\)http://www.diffusiontools.com/documentation/hysco.html
2.2. Existing susceptibility artifact correction methods

in the image volume so that the unwarped image is morphologically matched to the reference image. These methods have several variants based on the similarity measure between the EPI and reference images, e.g. the sum of squared differences [30], log-intensity metric [31], and mutual information [32, 33]. An advantage of this approach is that it does not require additional scans as the fieldmap-based methods do. However, methods in this class typically lack intensity distortion corrections and depend strongly on the constraints and parameters of the registration algorithms.

2.2.5 Hybrid SACs

Hybrid methods are the combination of two or more SAC approaches. A common hybrid approach is a combination between the PSF and fieldmap approaches. Xiang and Ye simplified the PSF method to form the fieldmap. The generated fieldmap is the phase difference between two distorted complex EPI images [37], thereby reducing the acquisition time. To reduce the acquisition time and correct both geometric and intensity distortion, Techavipoo et al. proposed a correction method that uses the algorithm in [23] to rewind the phase difference generated by Xiang and Ye.

Some hybrid methods incorporate the fieldmap and image registration approaches. For example, Gholipour et al. used the fieldmap as a guide for registering the distorted EPI images to the undistorted T$_{2w}$ structural MRI [39]. Daga et al. proposed a phase unwarping algorithm that produces the deformation field and uncertainty information [40]. The outputs of the phase unwarping algorithm are used as the initialization for the image registration from the distorted EPI image to the T$_{1w}$ image.

Other hybrid methods combined the fieldmap and reversed-PE approaches. Gelman et al. introduced a method combining the fieldmap and reversed-PE approaches to correct the geometric distortions and intensity distortions, respectively [41]. Bao et al. used the reversed-PE approach to eliminate the distortions
caused by the field inhomogeneities in the fieldmap, thereby increasing the accuracy of the correction method [42]. Another hybrid method, which combines the PSF and reversed-PE approach, was introduced by In et al..

2.3 Chapter summary

Susceptibility artifacts include intensity modulations and geometric distortions. These artifacts appear in MRI images due to the local field inhomogeneities which are mainly caused by the magnetic susceptibility differences of various images tissues, e.g. air versus bone and fat versus blood. Since the susceptibility artifacts result from the natural property of the imaged tissue, they are unavoidable in MRI images.

In EPI images, the susceptibility artifacts are significant because the main characteristic of the EPI pulse sequence is to image quickly. The geometric distortions occur mainly in the phase-encoding direction. They appear inversely in the image acquired with the reversed phase-encoding direction. Furthermore, the distortions are severe in high field strengths and high spatial resolutions.

There have been several methods for correcting the susceptibility artifacts in EPI images. However, some SAC methods require long scanning times or modifications of the acquisition sequence. Other methods, which correct the acquired-distorted images, require high computational resources; consequently, these SAC methods are unsuitable for time-sensitive applications, e.g. SAC on an MRI scanner. Among the SAC approaches, the reversed-PE based SAC is the most promising because of its efficiency. It is used to correct the fMRI and DWI data in the biggest MRI neuroimaging dataset - the HCP with 1200 subjects [48, 49].
Chapter 3

Susceptibility Artifact Correction using T1-weighted Regularization

Chapter contents

3.1 Introduction .................................................. 22
3.2 Hyper-elastic susceptibility artefact correction ............ 23
3.3 $T_1$w-guided reversed-PE SAC ............................ 27
   3.3.1 TISAC registration ................................. 27
   3.3.2 Optimization of hyper-parameters .................. 30
3.4 Experiments and results .................................. 32
   3.4.1 Data acquisition and preprocessing .................. 33
   3.4.2 Performance measures .............................. 36
   3.4.3 Experimental methods ............................. 38
   3.4.4 Analysis of the proposed method .................... 40
   3.4.5 Comparisons of SAC methods ....................... 40
   3.4.6 Discussion .......................................... 45
3.5 Chapter summary ........................................... 46
3.1 Introduction

This chapter aims to correct SAs in EPI-fMRI images, especially those with sub-millimeter resolutions. We propose to integrate a $T_{1w}$ structural image into a state-of-the-art susceptibility artifact correction iterative-optimization method, known as hyper-elastic susceptibility artifact correction (HySCO) [35]. The motivation is that the $T_{1w}$ image captures relatively well the shape and size of the tissue. It is widely considered a gold standard representation of a subject’s brain anatomy [50]. The $T_{1w}$ image can capture the high contrast between white-matter and gray-matter tissues [7]. Therefore, it is routinely acquired for every subject participating in fMRI studies, and it is readily available. We call the proposed method $T_{1w}$ guided Inverse phase encoding Susceptibility Artifact Correction, or TISAC.

The research contributions of this chapter can be highlighted as follows:

1. We introduce a new $T_{1w}$-based regularization term to the HySCO objective function to improve the quality of the corrected image with respect to the brain structure captured by the $T_{1w}$ image.

2. We select automatically the regularization parameters of the registration problem through a Bayesian optimization framework with a Gaussian process prior. Note that choosing the best regularization parameters is a critical step in solving the SAC optimization problem.

3. We evaluate the performance of the proposed method and compare it with existing SAC methods using two high-resolution EPI-fMRI datasets: one with an isotropic resolution of $1 \times 1 \times 1$ mm$^3$ acquired by a 3 Tesla (T) scanner, and the other with a resolution of $0.833 \times 0.833 \times 0.810$ mm$^3$ acquired by a

---

Chapter 3 has been published in:
3.2 Hyper-elastic susceptibility artefact correction

7T scanner. Experiment results show that the proposed method provides improved corrections that are well aligned to the $T_{1w}$ image.

The remainder of this chapter is organized as follows. Section 3.2 presents the general mathematical framework of the reversed-PE based correction method. Section 3.3 introduces our proposed method. Section 3.4 presents experiments and analysis of the proposed method and the related methods. Finally, Section 3.5 summarizes this chapter.

3.2 Hyper-elastic susceptibility artefact correction

This section presents the distortion model in the presence of the field inhomogeneity, and the hyper-elastic susceptibility artifact correction method.

First, let $E$ be the 3D ideal image, and $I$ be an acquired (distorted) image. As shown in [5, 11, 31], the distortion in the presence of field inhomogeneity $B$ in the image domain is modeled as

$$E = I(T) |J_T|,$$  \hspace{1cm} (3.1)

where $T$ is the non-rigid transformation operator of coordinates from image $E$ to image $I$, and $J_T$ is the Jacobian matrix of the transformation $T$. As shown in [11, 35], the transformation $T$ at any 3D point $p$ in $E$ can be written as $T: p \mapsto p + B(p)v$, where $v$ denotes the known distortion direction (i.e. the PE direction). In practice, the applied PE gradient is considered to be along the first dimension, hence $v = (1,0,0)$. Let $\partial_v(B(p))$ denote the directional derivative of field $B$ at point $p$ along the direction $v$. The Jacobian matrix of the transformation $T$ at point $p$ is

$$J_T(p) = \frac{\partial (p + B(p)v)}{\partial p} = \begin{bmatrix} 1 + \partial_v(B(p)) & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$  \hspace{1cm} (3.2)

The distortion model in Eq. (3.1) can be rewritten as

$$E(p) = I(p + B(p)v) \left[ 1 + \partial_v(B(p)) \right].$$  \hspace{1cm} (3.3)
3.2. Hyper-elastic susceptibility artefact correction

Here, the term $1 + \partial_v(B(p))$ denotes the intensity modulation. The term $p + B(p)v$ denotes the geometric displacement of the acquired image. In other words, point $p$ in ideal image $E$ is shifted to point $p + B(p)v$ in acquired image $I$. Since $B$ causes the voxel shifting in the acquired image, $B$ is called the *displacement field*, and $p + B(p)v$ is known as the deformation at point $p$. Fig. 3.1 illustrates the distortions caused by the displacement field. The ideal image in Fig. 3.1(a) under the displacement field in Fig. 3.1(b) is distorted, as shown in Fig. 3.1(c). It is worth noting that we work with 3D images; however, for simplicity, 2D images are presented throughout this chapter.

![Figure 3.1: A 2D illustration of the susceptibility-induced distortions. The displacement field is along the PE (horizontal) direction and is expressed in terms of the number of voxels shifted.](image)

Let $I_1$ and $I_2$ be two images of a subject in the same brain region, acquired using an identical sequence but with opposite blips. Let $B$ be the field inhomogeneity, and $v$ be the PE direction for image $I_1$. The field inhomogeneity and the PE direction for image $I_2$ are $B$ and $-v$, respectively. By applying the model in Eq. (3.3), the corrected images $E_1$ and $E_2$ can be described as

$$
\begin{align*}
E_1(p) &= I_1(p + B(p)v)[1 + \partial_v(B(p))], \\
E_2(p) &= I_2(p - B(p)v)[1 - \partial_v(B(p))].
\end{align*}
$$

For notational simplicity, hereinafter $X_p$ will refer to the intensity of image $X$ at location $p$. 

24
3.2. Hyper-elastic susceptibility artefact correction

Next, recall that the reversed-PE approach estimates the displacement field $B$ based on two images $I_1$ and $I_2$ acquired using an identical sequence but with opposite blips. Field $B$ is estimated such that two corrected images $E_1$ and $E_2$ are as similar as possible. The estimated $B$ is then used to unwarp the distorted images $I_1$ and $I_2$ based on Eq. (3.4).

The hyper-elastic susceptibility artifact correction method proposed by Ruthotto et al. [35] uses the reversed-PE approach to correct SAs. To estimate $B$, Ruthotto et al. minimized the SSD-based dissimilarity between unwarped images $E_1$ and $E_2$ [11, 35]:

$$D(I_1, I_2, B) = D(E_1, E_2) = \frac{1}{2} \int_{\Omega} (E_{1p} - E_{2p})^2 dp. \quad (3.5)$$

Finding $B$ by minimizing the distance function $D(I_1, I_2, B)$ is categorized as an ill-posed problem [11, 35]. Thus, prior knowledge about the smoothness of the displacement field and invertibility of the geometrical transformation was used to regularize $B$ [35]. To enforce the smoothness of the displacement field, a Tikhonov ($\mathcal{L}_2$) regularizer $S_{\text{diff}}$ was integrated into the objective function [11]:

$$S_{\text{diff}}(B) = \int_{\Omega} \| \nabla B_p \|^2 dp. \quad (3.6)$$

To satisfy the invertibility of the transformation, the Jacobian matrix of the geometric transformation in Eq. (3.4) must be invertible. In other words, Jacobian determinants must be positive for all $p \in \Omega$. [5] demonstrated that this constraint could be expressed as $-1 \leq \partial v(B_p) \leq 1$, for all $p \in \Omega$.

Ruthotto et al. inspired by the control of volumetric change in hyper-elasticity [51], introduced an additional non-linear term $S_{\text{hyper}}$ to the objective function:

$$S_{\text{hyper}}(B) = \int_{\Omega} \phi(\partial v(B_p)) dp, \quad \text{with} \quad \phi(z) = \frac{z^4}{1-z^2}. \quad (3.7)$$
3.2. Hyper-elastic susceptibility artefact correction

Collectively, Ruthotto et al. proposed the objective function:

\[ J(B) = D(I_1,I_2,B) + \alpha S_{\text{diff}}(B) + \beta S_{\text{hyper}}(B), \quad \text{s.t.} \quad |\partial v(B_p)| \leq 1. \quad (3.8) \]

The positive and user-defined regularization parameters \( \alpha \) and \( \beta \) represent the trade-off between the smoothness and the elasticity of the displacement field \( B \).

The HySCO method estimates \( B \) by minimizing the objective function \( J(B) \) in Eq. (3.8), then generates the output (corrected) images using Eq. (3.4). HySCO can provide output images with high similarity; however, these images are blurry, and they may not align well with the actual brain structure. For example, Fig. 3.2(a) shows the estimated deformation grid\(^b\) by HySCO, and Fig. 3.2(b) shows the output image. The output image contains blur trails (e.g. areas denoted by the red arrows), which are caused by over-deformation in the estimated field \( B \). This over-deformation could be reduced by imposing an independent constraint related to the brain structure.

\(^b\)The deformation grid is the sum of the regular grid and the displacement field.
3.3 \textbf{T}_{1w}-\text{guided reversed-PE SAC}

The \text{T}_{1w} structural image, acquired using MPRAGE sequence [52] or MP2RAGE sequence [53], is widely considered to reflect the anatomical structure of the brain, especially in the fMRI study [50]. In this chapter, we propose using the \text{T}_{1w} image to guide the susceptibility artifact correction of high spatial resolution EPI-fMRI images. In the proposed approach, the \text{T}_{1w} image is used for two purposes: (i) introducing an additional regularization term for the new objective function, and (ii) selecting the three regularization parameters of the objective function.

3.3.1 \textbf{TISAC registration}

The reversed-PE correction problem integrated with a \text{T}_{1w} structural image can be formulated as finding the displacement field $B$ such that the corrected (unwarped) images $E_1$ and $E_2$ satisfy two criteria: (i) be as similar to each other as possible, and (ii) align well with the structural information provided by the \text{T}_{1w} image. \cite{35} proposed the objective function in Eq. (3.8), which satisfies the first criterion. We introduce a \text{T}_{1w}-\text{guided regularization term to address the second criterion. More precisely, the regularization term measures the dissimilarity between the multi-modal images, \text{i.e.} \text{T}_{1w} and EPI-fMRI. Conceptually, minimizing the proposed objective function is equivalent to minimizing the dissimilarity between the corrected EPI-fMRI images and the dissimilarity between corrected images and the \text{T}_{1w} image. This subsection is designed to provide an accessible mathematical description of the proposed method.

The proposed regularization term is based on the normalized gradient field (NGF), which has been proven to be well-suited for the multi-modal registration problem [54]. The NGF measure, at any point in an image, reveals the intensity change and its direction. Let $\nabla X_p$ be the gradient at point $p$ of image $X$, and $\epsilon$ be a user-defined parameter. As shown in [54], the NGF measure at point $p$ is defined
3.3. $T_{1w}$-guided reversed-PE SAC

as

$$\tilde{\nabla}X_p = \frac{\nabla X_p}{\sqrt{||\nabla X_p||^2 + \epsilon^2}}. \quad (3.9)$$

The difference between two images $X$ and $Y$ can be measured using the angles formed by NGF vectors at all points in the image domain. Accordingly, the NGF-based distance between two images $X$ and $Y$ is defined as

$$D_{NGF}(X,Y) = \frac{1}{2} \int_\Omega \left[ 1 - \langle \tilde{\nabla} X_p, \tilde{\nabla} Y_p \rangle^2 \right] d\mathbf{p}, \quad (3.10)$$

where $\langle \cdot, \cdot \rangle$ denotes the dot-product operator. The value of $D_{NGF}(X,Y)$ is positive. The smaller the value of $D_{NGF}(X,Y)$ is, the more similar are the two images.

Let $A$ denote the $T_{1w}$ image. We introduce the $T_{1w}$-guided regularization term as the sum of the NGF-based distances of image $A$ to each unwarped image of $I_1$ and $I_2$ under the displacement field $B$

$$D_A(I_1, I_2, B, A) = D_{NGF}(A, E_1) + D_{NGF}(A, E_2)$$

$$= \frac{1}{2} \int_\Omega \left[ 1 - \langle \tilde{\nabla} A_p, \tilde{\nabla} E_{1p} \rangle^2 \right] + \left[ 1 - \langle \tilde{\nabla} A_p, \tilde{\nabla} E_{2p} \rangle^2 \right] d\mathbf{p}. \quad (3.11)$$

To summarize, we introduce a new objective function:

$$J(B) = D(I_1, I_2, B) + \alpha S_{\text{diff}}(B) + \beta S_{\text{hyper}}(B) + \gamma D_A(I_1, I_2, B, A) \quad (3.12)$$

s.t. \ $|\partial_v(B_p)| \leq 1$ \ for all $\mathbf{p} \in \Omega$.

The displacement field is found by minimizing $J(B)$ in Eq. (3.12). The positive and user-defined regularization parameters $\alpha$, $\beta$, and $\gamma$ represent the trade-off between the similarity of the corrected images, the smoothness of $B$, the elasticity of the displacement, and the similarity to the $T_{1w}$ image of corrected images.

In this chapter, the Gauss-Newton method is used for minimization. This method starts with an initial guess of $B$, e.g. $B^{(0)} = 0$. The next estimate of $B$ is computed iteratively as

$$B^{(k+1)} = B^{(k)} - \lambda^{(k)} G^{(k)} (H^{(k)})^{-1}, \quad \lambda^{(k)} > 0, \quad (3.13)$$
3.3. $T_{1w}$-guided reversed-PE SAC

![Diagram of coarse-to-fine optimization scheme]

Figure 3.3: The block diagram of the coarse-to-fine optimization scheme. The displacement field is estimated at each level of data representation.

where superscript $k$ is the iteration number, $\lambda^{(k)}$ is the learning rate, and $G^{(k)}$ and $H^{(k)}$ are the approximate gradient and Hessian of the objective function $J$, respectively.

A small learning rate leads to slow convergence, while a large one may lead to invalid $B^{(k+1)}$. Therefore, to select a suitable learning rate, we find the maximum $\lambda^{(k)}$ that produces $B^{(k+1)}$ meeting the constraint in (3.12) [55]. This is done by applying the backtracking line search [56].

To avoid local minima and to accelerate the convergence, the Gauss-Newton method is integrated with the coarse-to-fine approach (see Fig. 3.3). This approach first represents images with multiple resolution levels. The image representation at a coarser level is obtained simply by averaging over adjacent cells. Next, the displacement field in the coarsest level is estimated by minimizing the objective function in (3.12) using the image representation at this level. The estimated displacement field at the coarser level is interpolated. The interpolated result
is considered the initial guess for the optimizer at a finer level. The process of interpolation and estimation is repeated until the displacement field at the finest level is obtained. Finally, the corrected images are obtained by unwarping the distorted images with the estimated field $B$, as shown in Eq. (3.4). This coarse-to-fine optimization approach is summarized in Algorithm 1.

**Algorithm 1** Coarse-to-fine Gauss-Newton for SAC

**Input:** $I_1, I_2$: reversed-PE EPI-fMRI images,  
$A$: $T_{1w}$ image corresponding to fMRI images,  
$l_{\text{min}}, l_{\text{max}}$: min, max level of data representation.

**Output:** Corrected images $E_1$ and $E_2$.

1. Derive the multilevel image representation;
2. $B_{l_{\text{min}}-1} \leftarrow 0$;
3. for $l = l_{\text{min}} : l_{\text{max}}$ do
4. Interpolate $B^{(0)}_l$ from $B_{l-1}$: $B^{(0)}_l \leftarrow \text{inter}(B_{l-1})$;
5. $k \leftarrow 0$;
6. Compute the objective function as in Eq. (3.12):
   $[J, C^{(k)}, H^{(k)}] \leftarrow \text{obj_fnct}(I_1, I_2, B^{(k)}_l, A, l)$;
7. while not converged do
8. Compute the new $B$ via backtracking line search:
   $B^{(k+1)}_l \leftarrow \text{backtrack_search}(B^{(k)}_l, G^{(k)}, H^{(k)})$;
9. Increment $k$: $k \leftarrow k + 1$;
10. Compute the objective function as in Eq. (3.12):
    $[J, C^{(k)}, H^{(k)}] \leftarrow \text{obj_fnct}(I_1, I_2, B^{(k)}_l, A, l)$;
11. end while
12. $B_l \leftarrow B^{(k)}_l$;
13. end for
14. Unwarp $I_1$ and $I_2$ using Eq. (3.4)
   $E_1 \leftarrow \text{unwarp}(I_1, B_{l_{\text{max}}})$;
   $E_2 \leftarrow \text{unwarp}(I_2, B_{l_{\text{max}}})$;

### 3.3.2 Optimization of hyper-parameters

In the reversed-PE SAC, the choice of the regularization parameters (hyper-parameters) is crucial. Here, we propose a method to select the most suitable regularization parameters for the SAC problem. The proposed hyper-parameters optimization method is based on the Bayesian optimization (BO) with a Gaussian process (GP) prior.
The hyper-parameter optimization is performed by minimizing an error function \( f(x) \) of the given SAC method over a dataset \( D \), where \( x \) is a vector of hyper-parameters. The error function here is defined as the sum of the dissimilarity measure \( M \) between the \( T_{1w} \) image and the corrected fMRI images for the dataset \( D \). In this chapter, the MIND-based measure is used (refer to Appendix B for a description of the MIND measure). The mathematical equation of the loss function is:

\[
 f(x) = L(S_x, D) = \frac{1}{|D|} \sum_{I_{1i}, I_{2i}, A_i \in D} M(S_x(I_{1i}, I_{2i}), A_i) 
 = \frac{1}{|D|} \sum_{I_{1i}, I_{2i}, A_i \in D} M(E_{1i}^{S_x}, A_i) + M(E_{2i}^{S_x}, A_i),
\] (3.14)

where \( S_x(I_{1i}, I_{2i}) \) represents the corrected images \( E_{1i}^{S_x} \) and \( E_{2i}^{S_x} \) for the inputs \( I_{1i} \) and \( I_{2i} \), by applying the SAC method \( S \) with hyper-parameters \( x \). We select the hyper-parameters which give minimum error function. In other words, finding the hyper-parameters is to minimize the error function Eq. (3.14). Since the error function of the hyper-parameters is computationally expensive, and its distribution is unknown, the hyper-parameter optimization problem is challenging.

BO is a powerful technique for finding extrema of an objective function that has no closed-form expression or is computationally intensive to evaluate [57, 58, 59]. The BO algorithm uses previous observations, which are pairs of \( \{x, f(x)\} \), to determine what is the next optimal point for sampling the error function.

To be specific, the BO algorithm first computes the posterior expectation of what the function \( f \) looks like based on its previous observations. This step is done by first considering that the distribution of \( f(x) \) is a normal likelihood with noise. The error function \( f \) then can be considered a Gaussian process, which is specified by the mean \( \mu \) and variance \( \sigma \) of a normal distribution over possible values of \( f(x) \). The means and covariances allow us to update our belief of what the function \( f \) looks like. They can be obtained by fitting the GP to a given set of observations \( \mathcal{H} = \{(x_1, f(x_1)), (x_2, f(x_2)), \ldots, (x_n, f(x_n))\} \).
Next, a new point is selected to sample the function $f$ so that it provides a higher value of $f$ or is in the unexplored region. As shown in [58], the point can be found by maximizing the expected improvement function, which is defined as

$$
\Psi(x) = \begin{cases} 
\left[ \mu(x) - f(x^*) \right] \Phi(z) + \sigma(x) \phi(z) & \text{if } \sigma(x) > 0 \\
0 & \text{if } \sigma(x) = 0,
\end{cases}
$$

(3.15)

where $x^*$ is the current optimal hyper-parameter point, $\mu(x)$ and $\sigma(x)$ are the estimated mean and variance of function $f$ at $x$ in the previous step, $z = \frac{\mu(x) - f(x^*)}{\sigma(x)}$, $\Phi(z)$ is the cumulative distribution, and $\phi(z)$ is probability density function of the standard normal distribution.

**Algorithm 2** Hyper-parameters optimization algorithm.

**Input:** $D$: dataset.

**Output:** $x^*$: optimal hyper-parameters.

1: $\mathcal{H} \leftarrow \emptyset$;
2: while not converged do
3: Fit GP on the observation set: $[\mu, \sigma] \leftarrow \text{GP}(\mathcal{H})$;
4: Choose the next point for sampling: $\hat{x} \leftarrow \text{argmax}_x \Psi(x|\mu, \sigma)$;
5: Compute the error function at $\hat{x}$: $f(\hat{x}) \leftarrow L(S_{\hat{x}}, D)$;
6: Update the observation set: $\mathcal{H} \leftarrow \mathcal{H} \cup (\hat{x}, f(\hat{x}))$;
7: Increment $k$: $k \leftarrow k + 1$;
8: end while
9: $x^* \leftarrow \text{argmin}_{x \in \mathcal{H}} f(x)$

The new point obtained by maximizing the expected improvement function $\Psi(x)$ is admitted to the observation set. The procedure of fitting GP and finding the sampling point is repeated until the convergence criterion is met. Algorithm 2 shows how the Bayesian optimization problem for automatically selecting the hyper-parameters can be solved iteratively.

### 3.4 Experiments and results

This section presents the experiments and analysis of the proposed method. Section 3.4.1 describes data acquisition and preprocessing, and Section 3.4.2 presents
3.4. Experiments and results

the evaluation measures. Section 3.4.3 shows the experimental methods, and Section 3.4.4 presents an analysis of the proposed method. Section 3.4.5 compares TISAC with the state-of-the-art SAC methods. Finally, Section 3.4.6 discusses the experimental results.

3.4.1 Data acquisition and preprocessing

Two EPI-fMRI datasets of the occipital cortex were used to evaluate the performance of SAC methods. The first dataset had three subjects and was acquired using a 3T scanner with an isotropic resolution of $1 \times 1 \times 1 \text{mm}^3$. The second dataset had three subjects and was acquired using a 7T scanner with a resolution of $0.833 \times 0.833 \times 0.810 \text{mm}^3$. A brief summary of these datasets is presented in Table 3.1. The datasets were acquired with the written informed consent from all participating subjects, in accordance to the Human Ethics Committees requirements at the University of Queensland, and the Australian National Health and Medical Research Council’s guidelines.

Table 3.1: A summary of the datasets used in the experiments.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>No. subjs.</th>
<th>Gender dist.</th>
<th>Age avg.</th>
<th>Volume size</th>
<th>Resolution (mm$^3$)</th>
<th>Acquisition sequence</th>
<th>Field strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>3T</td>
<td>3</td>
<td>1 female, 2 males</td>
<td>23.67</td>
<td>$192 \times 144 \times 36$</td>
<td>$1 \times 1 \times 1$</td>
<td>2D single-shot GRE-EPI</td>
<td>3T</td>
</tr>
<tr>
<td>7T</td>
<td>3</td>
<td>3 males</td>
<td>35.33</td>
<td>$192 \times 192 \times 48$</td>
<td>$0.833 \times 0.833 \times 0.810$</td>
<td>3D GRE-EPI (WIP1080)</td>
<td>7T</td>
</tr>
</tbody>
</table>

The 3T dataset from three healthy subjects was acquired using a Siemens 3T MAGNETOM PRISMA with a 64-channel head coil and a 2D single-shot gradient-echo EPI sequence. Ascending and interleaved coronal slices were acquired with a repetition time (TR) of 3000 milliseconds (ms), which is also the volume repetition time, TE of 30 ms, a flip angle of 90 degrees, and an image size of $192 \times 144 \times 36$. The field of view (FOV) was $144 \text{mm} \times 192 \text{mm}$.

The 7T dataset from three healthy subjects was acquired using a Siemens 7T MAGNETOM whole-body research scanner with a 32-channel head coil (Nova
3.4. Experiments and results

Medical, Wilmington, US) and a 3D EPI sequence WIP1080 (Poser et al., 2010) [60]. The sequence used a blipped CAIPIRINHA [61, 62], implementation [63] with the following parameters: TE of 30 ms, TR of 83 ms, volume repetition of 1992 ms, flip angle of 17 degrees, echo spacing of 1 ms, FOV of 160 mm × 160 mm, matrix size of 192 × 192 × 48. The image acquisition was accelerated by a factor of 2 in-plane and by a factor of 2 in the slice-encoding direction with a CAIPI-shift of 1. This results in a total acceleration factor of 4. The image reconstruction was done by using the GRAPPA pipeline [64], as provided by the vendor.

Figure 3.4 shows the three different orientation views (coronal, sagittal, and axial) of 7T reversed-PE EPI images (pink) overlaying on the T$_{1w}$ image (green). The figure demonstrates that the misalignment of EPI to the T$_{1w}$ image occurs mainly in one spatial direction (left-to-right).

![Figure 3.4: Distorted EPI slices overlaying on a T$_{1w}$ image with three different orientation views. The blue lines (cross-hairs) indicate the intersection point of the three views. See the electronic color images.](image)

Functional MRI data were acquired while subjects were presented with retinotopic mapping stimuli. In the 3T dataset, stimuli consisted of drifting bars, expanding rings, rotating bowties, and flashing full-field (see Fig. 3.5). Each subject took part in two scanning sessions; in each session, subjects viewed visual stimuli while scanning using either left-to-right or right-to-left blips, such that each
blip accounted for half the scans. This resulted in pairs of scans with reversed patterns of distortions in the PE direction. In the 7T dataset, only the rotating bowtie stimulus was used. In each subject, two scans (with 183 or 187 volumes each) were collected with LR blip, and two short 20 s measurements with ten repeated EPI volumes were collected with the inverse blip, one at the beginning of the experimental runs and one at the end.

Figure 3.5: Examples of visual stimuli presented to the subjects during scanning.

For each subject in the 3T dataset, a $T_{1w}$ image of the entire-brain was acquired using the 3D GRE-MRI sequence, with cubic voxels of 0.75 mm edge length. The $T_{1w}$ image was then upsampling into an image with a resolution of $0.5 \times 0.5 \times 0.5$ mm$^3$. For each subject in the 7T dataset, a whole-brain anatomical image was collected using an MP2RAGE sequence WIP900b17a [53], with a resolution of $0.500 \times 0.533 \times 0.533$ mm$^3$.

In the first preprocessing step, all fMRI images were motion-corrected using tools in SPM12 [65]. The 3T dataset was also slice scan time corrected. Hereinafter, the data without motion and slice scan time corrections are referred to as original data; and the preprocessed images without SAC are referred to as uncorrected data. A $T_{1w}$ alignment image of each subject was created by aligning the $T_{1w}$ image to an average of two oppositely-distorted images of the subject, through SPM’s co-registration procedure [66].
3.4.2 Performance measures

We quantitatively evaluate the corrected images in three aspects: geometric correction, blurriness, and the suitability for BOLD analysis. The various performance measures are described in this subsection.

Structural similarity measures are used to evaluate how well the corrected fMRI image matches the brain structural given by the $T_{1w}$ image. Here, we used the mutual information (MI) to compute the similarity between the fMRI images and the $T_{1w}$ image [67]. A smaller value of MI indicates less similarity between the functional and structural $T_{1w}$ images.

The percentage of activated voxel evaluates both the geometric accuracy and the suitability for subsequent BOLD analysis. The reason is that the BOLD response is localized in gray matter and to a certain degree in the cerebrospinal fluid (CSF) more for 3T and less for 7T data, but not in white matter. Distortions of fMRI images result in some significantly modulated voxels being mislocated in white matter of the $T_{1w}$ image. Here, we employ correlation analysis, a common and robust method for analyzing phase-encoded retinotopic mapping data. This analysis provides a phase-map\(^c\) of the BOLD responses [68, 69]. Fig. 3.6 shows an example of a phase-map obtained by correlation analysis of an uncorrected fMRI scan. Voxels with supra-threshold response are marked in color, where the color depicts the phase (delay) of the response, not the strength of the activation. In the given example, there are many activated voxels located in white matter, indicating that they are displaced by distortions.

In this chapter, we introduce a measure to evaluate the geometric correction in the corrected images using the percentage of activated voxels in gray matter and white matter. The reasons of measuring the percentage of activated voxels in white matter are: (i) white matter is surrounded by gray matter; (ii) there is a large

\(^c\)The term phase-map refers to the use in phase-encoded retinotopic mapping, which is provided by an FFT-based analysis procedure of BOLD time courses. It is different from “phase” in the phase-encoding direction derived from k-space in MRI acquisition.
3.4. Experiments and results

Figure 3.6: An example of a phase-map in the coronal plane. The red line marks the outer boundary of white matter. Note that the color coding represents the position in the visual field (see the color wheel), not the strength of responses, as typical in phase-encoded retinotopic mapping.

number of activated voxels aligned to white matter; and (iii) it is easy to obtain an accurate and reliable segmentation of white matter from the T_{1w} image. The percentage of activated voxels in CSF is not considered as it is not diagnostic for geometric accuracy. A higher percentage of activated voxels in gray matter and a lower percentage of activated voxels in white matter indicates a better alignment of the fMRI images to the T_{1w} image.

Blurriness measure is used to evaluate how blurry the image is. Introducing blur to high spatial resolution fMRI data is typically undesirable [7, 70], as it negates the often considerable effort to achieve high spatial resolution. To measure blurriness, we extended the measure proposed by [71] for 2D images, to work for 3D images. This measure reflects the intensity variation of an image with respect to that of the low-pass filtered image.

The normalized intensity variation of image $I$ in the $i^{th}$ direction is defined as

$$V_i(I) = \frac{\sum_{p \in \Omega} \max\{0, \partial_i(I_p) - \partial_i(I_p \otimes h_i)\}}{\sum_{p \in \Omega} \partial_i(I_p)},$$

(3.16)

where $h_i$ is a low-pass filter, and $\partial_i(I_p)$ is the partial derivative at point $p$ in the $i^{th}$
3.4. Experiments and results

direction. The blurriness measure for a 3D image $I$ is the sum of the normalized intensity in three directions

$$M_b(I) = \frac{1}{3} \sum_{i=1,2,3} [1 - V_i(I)].$$  \hspace{1cm} (3.17)

An image with a higher value of $M_b$ is more blurred than the one with a lower value. Fig. 3.7 shows an example of the blurriness measure.

![Figure 3.7: Blurriness measurements of an original 3D T1w image and two blurred images, which are produced by two Gaussian smoothing filters with different standard deviations $\sigma$.](image)

**Suitability for BOLD analysis** measures undesired changes in the BOLD responses. For this, we estimate the cumulative distribution function (CDF) of the phase-map values in every slice. The suitability for BOLD analysis is defined as the difference between the CDFs of corrected and uncorrected data. It is measured by the normalized cross-correlation (NCC) function. The range of NCC is in [0, 1]. A small value of NCC indicates a significant change of the BOLD responses between the corrected fMRI images and uncorrected images and vice versa.

3.4.3 Experimental methods

Scans of inverse blips were first paired together. A mean image over time of each scan was then generated. The mean images of each scan pair were processed by the SAC methods to estimate the displacement field. The estimated displacement field was then used to unwarp all volumes in the scan pair. TISAC and HySCO use the same framework implemented in MATLAB to unwarp the distorted im-
ages, while TOPUP uses another framework implemented in C programming language. However, these unwarping frameworks are all based on the cubic spline interpolation.

We evaluated the sub-components of TISAC, which are $T_{1w}$-based registration ($T_{1w}$R), and Bayesian optimization (BO). The tested configurations include: (i) $T_{1w}$R only; and (ii) $T_{1w}$R with BO (i.e. the complete TISAC). For the configuration of $T_{1w}$R only, the regularization parameters were selected as $\alpha = 30$, $\beta = 50$, and $\gamma = 75000$.

We further compared the proposed TISAC method with two state-of-the-art SAC methods: HySCO (from the SPM12 toolbox version r7219) and TOPUP (from the FSL package version 5.0.9). For each pair of reversed-PE scans, the displacement field was estimated using two mean images of these scans and then used to unwarp the distorted images from these scans. The regularization parameters of TISAC were selected automatically by applying the BO technique, while these parameters in HySCO were set as $\alpha = 50$ and $\beta = 10$ as suggested in [35]. The regularization parameters of TOPUP were selected as indicated in the preprocessing pipeline for the HCP [49].

We also assessed the time complexity of three SAC methods by recording their execution time with inputs as pairs of mean images. All timing results were collected on a Linux workstation with an Intel Xeon Processor E3-128V2 3.6 GHz and 32 GB RAM.

We evaluated the statistical significance of the measures using two-sample $t$-tests with the Bonferroni correction. This approach is simple and robust against false positives. The one-way analysis of variance (ANOVA) was used to test the differences in blurriness, MI measure, and percentage of activated voxels among SAC methods. All the tests (two-sample $t$-tests and ANOVAs) were implemented using MATLAB. The $t$-test produces a $p$-value, which is used to evaluate the statistical significance of the test (default significance level is 0.05). A smaller $p$-value indicates stronger evidence against the null hypothesis ($H_0$). A $p$-value...
3.4. Experiments and results

less than 0.05 means the null hypothesis is rejected at a confidence level of 95%. A \( p \)-value greater than 0.05 indicates weak evidence against the null hypothesis.

### 3.4.4 Analysis of the proposed method

We investigated whether the BO technique improves the \( T_{1w} \)-based registration scheme. Table 3.2 shows the similarity measures and execution time of the two TISAC settings. It appears that all tested configurations provided corrected images with comparable quality, \textit{i.e.} similarity to the \( T_{1w} \) structural image. However, using the BO technique led to a faster run time than when it was not used.

**Table 3.2:** Similarity measures and execution times (in second) of TISAC configurations: with and without using the BO technique. Methods using BO do not include time for estimating the hyper-parameters.

<table>
<thead>
<tr>
<th>Measures</th>
<th>( T_{1w} )R (mean ± std)</th>
<th>( T_{1w} )R + BO (mean ± std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIND</td>
<td>0.28 ± 0.02</td>
<td>0.27 ± 0.02</td>
</tr>
<tr>
<td>Execution</td>
<td>25.60 ± 9.33</td>
<td>19.71 ± 8.36</td>
</tr>
</tbody>
</table>

### 3.4.5 Comparisons of SAC methods

First, we investigated the time complexity of the proposed method. Table 3.3 shows the processing time comparison of three SAC methods: TISAC, TOPUP, and HySCO. The results indicate that the proposed TISAC is significantly faster than TOPUP and HySCO (\( p \)-values < 0.05). TISAC is approximately 24.8 times faster than TOPUP, and 1.4 times faster than HySCO.

**Table 3.3:** Comparison with other SAC methods in terms of execution time (in second).

<table>
<thead>
<tr>
<th>Datasets</th>
<th>TOPUP mean ± std, ( p )-value</th>
<th>HySCO mean ± std, ( p )-value</th>
<th>TISAC mean ± std</th>
</tr>
</thead>
<tbody>
<tr>
<td>3T</td>
<td>399.87 ± 6.52, 0.000</td>
<td>21.19 ± 5.38, 0.000</td>
<td>14.88 ± 1.78</td>
</tr>
<tr>
<td>7T</td>
<td>741.33 ± 6.04, 0.000</td>
<td>43.88 ± 12.03, 0.032</td>
<td>32.58 ± 3.80</td>
</tr>
</tbody>
</table>

The \( t \)-test \( \mathcal{H}_0: t_{\text{TISAC}} \geq t_{\text{other}} \). The \( p \)-values are Bonferroni corrected.

Second, we visually assessed the quality of corrected images generated by the three SAC methods. Fig. 3.8 shows the uncorrected and the corresponding corrected images of these SAC methods for two subjects, one in the 3T dataset. 

---

40
3.4. Experiments and results

(top row) and the other in the 7T dataset (bottom row). The corresponding $T_{1w}$ images are presented in the right-most column. The figure shows that all tested SAC methods decreased SA distortions noticeably. In both datasets, the TISAC method produced sharp images with clearly visible tissue interfaces, especially near the brain-air interface. In comparison, TOPUP produced low contrast images, and HySCO produced images with artifacts in the brain-air interface (see cyan arrows).

![Figure 3.8: Uncorrected images and their corrected versions created using three SAC methods and corresponding $T_{1w}$ images. Top row: images of a subject in the 3T dataset. Bottom row: images of a subject in the 7T dataset. The arrows point to the artifacts produced by HySCO.](image)

Third, we analyzed the level of blurriness that each SAC method produces. Fig. 3.9 shows the cumulative distribution of the blurriness measurements of the two datasets for five cases: uncorrected, TOPUP, HySCO, TISAC, and Gaussian filtering of the uncorrected data with a standard deviation of $\sigma = 0.3$. Note that the SA-uncorrected data were obtained by applying the motion correction (MOCO) to the original data. The SAC methods were applied after the MOCO. We observed that the TISAC produced corrected data with the least blur among three SAC methods. Compared to HySCO, TISAC added significantly less blur on both datasets ($p$-value = 0.000, see Table 3.4). Compared to TOPUP, TISAC added significantly less blur on the 3T dataset ($p$-value = 0.000), and slightly less blur on the 7T dataset ($p$-value = 0.078). All three SAC methods added blur into the uncorrected data. This observation was confirmed by an one-way ANOVA for each
3.4. Experiments and results

dataset (3T dataset: $F_{3,24,956} = 2889.36$, $p$-value = 0.000; 7T dataset: $F_{3,6,664} = 129.59$, $p$-value = 0.000).

![Cumulative distributions of the blurriness measurements of the two datasets for five cases: uncorrected, TOPUP, HySCO, TISAC, and Gaussian filtering of the uncorrected data.](image)

**Figure 3.9:** Cumulative distributions of the blurriness measurements of the two datasets for five cases: uncorrected, TOPUP, HySCO, TISAC, and Gaussian filtering of the uncorrected data.

**Table 3.4:** Comparison with other SAC methods in terms of the blurriness introduced.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>TOPUP mean ± std, $p$-value</th>
<th>HySCO mean ± std, $p$-value</th>
<th>TISAC mean ± std</th>
</tr>
</thead>
<tbody>
<tr>
<td>3T</td>
<td>0.271 ± 0.006, <strong>0.000</strong></td>
<td>0.277 ± 0.008, <strong>0.000</strong></td>
<td>0.269 ± 0.006</td>
</tr>
<tr>
<td>7T</td>
<td>0.232 ± 0.012, <strong>0.078</strong></td>
<td>0.236 ± 0.014, <strong>0.000</strong></td>
<td>0.231 ± 0.011</td>
</tr>
</tbody>
</table>

*The $t$-test $H_0: b_{\text{TISAC}} \geq b_{\text{other}}$. The $p$-values are Bonferroni corrected.*

The proposed TISAC added a similar amount of blur to the MOCO corrected data as did a Gaussian filter with a standard deviation of $\sigma = 0.3$. From a related experiment, we observed that motion correction added a similar amount of blur to the original data as did a Gaussian filter with a standard deviation of $\sigma = 0.35$. Therefore, using the standard deviations of the Gaussian filter as references, we can conclude that TISAC adds less blur to the MOCO corrected images than what motion correction adds to the original images.

Fourth, we computed the structural similarity measures for more quantitative evaluation. Fig. 3.10 shows box-plots of MI coefficients of the two datasets. For the 3T dataset, the MI coefficients shows a small change between corrected and uncorrected images. An one-way ANOVA showed a significant main effect ($F_{3,49,284} = 5.19$, $p$-value = 0.001). However, the post-hoc tests revealed that the
MI coefficients for TOPUP are significantly larger than all others. The post-hoc tests also showed no difference between the uncorrected, HySCO, and TISAC. For the 7T dataset, all three SAC methods improved the MI coefficients (one-way ANOVA: $F_{3,11397} = 3.08$, $p$-value = 0.026), and there was no significance difference between SAC methods (one-way ANOVA: $F_{2,4998} = 1.68$, $p$-value = 0.187).

![Box-plots of the MI coefficients between structural T1w and fMRI images.](image1)

**Figure 3.10:** Box-plots of the MI coefficients between structural T1w and fMRI images.

![Phase-maps projected onto the T1w image of uncorrected and corrected data in the 7T data of Subject 1.](image2)

**Figure 3.11:** Phase-maps projected onto the T1w image of uncorrected and corrected data in the 7T data of Subject 1. Top row: phase-maps in the coronal view. Bottom row: phase-maps in the axial view. The arrows point to the areas with large distortions. See the electronic color images.

Fifth, we tested if TISAC improves the accuracy of geometric correction via the BOLD localization in gray matter and white matter. Note that, a high PAV score is desirable in gray matter, whereas a low PAV score is desirable in white matter. Figs. 3.11 to 3.13 show phase-maps of uncorrected and three SAC corrected data of a subject in the 7T dataset, with the coronal and axial views. Visual inspection reveals that geometric distortions are smaller in the 3T datasets than in the 7T datasets, where uncorrected 7T data exhibit a clear misalignment between...
3.4. Experiments and results

Figure 3.12: Phase-maps projected onto the $T_{1w}$ image of uncorrected and corrected data in the 3T data of Subject 4. See the electronic color images.

Figure 3.13: Phase-maps projected onto the $T_{1w}$ image of uncorrected and corrected data in the 7T data of Subject 3. See the electronic color images.

activated voxels and gray matter. The maximum misalignment is 5 pixels (equivalent to about 4.16 mm, see the arrows on the phase-maps of uncorrected data in Fig. 3.11). Also, visual inspection suggests that TISAC correction produces better alignment than HySCO, and slightly better alignment than TOPUP.

Figure 3.14: Mean percentage of activated voxels in gray matter (GM) and white matter (WM). The error bar shows the corresponding standard deviation of the percentage.
3.4. Experiments and results

To quantify this, we calculated the percentage of activated voxels (PAV). Fig. 3.14 shows the PAV measures in gray matter and white matter of the 3T and 7T datasets. For the 3T dataset, the SAC methods did not improve the PAV in gray matter (one-way ANOVA: $F_{2,4998} = 0.84$, $p$-value = 0.473). They also did not decrease the PAV in white matter (one-way ANOVA: $F_{2,4998} = 0.77$, $p$-value = 0.511).

For the 7T dataset, the PAV measures show a strong improvement of TISAC corrected (see Fig. 3.14), i.e. the PAV measure in GM of TISAC (mean = 61.22, std = 1.30) is greater than the PAV measure in GM of HySCO (mean = 56.63, std = 1.87), TOPUP (mean = 57.23, std = 1.12), and uncorrected data (mean = 54.66, std = 5.34). However, as there is only a small number of samples, we did not perform a statistical test.

Table 3.5: Change of the BOLD responses after SACs over scans.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>TOPUP mean ± std, $p$-value</th>
<th>HySCO mean ± std, $p$-value</th>
<th>TISAC mean ± std</th>
</tr>
</thead>
<tbody>
<tr>
<td>3T</td>
<td>0.971 ± 0.089, 0.188</td>
<td>0.964 ± 0.078, 0.739</td>
<td>0.962 ± 0.089</td>
</tr>
<tr>
<td>7T</td>
<td>0.998 ± 0.009, 0.525</td>
<td>0.999 ± 0.004, 0.968</td>
<td>0.999 ± 0.004</td>
</tr>
</tbody>
</table>

The $t$-test $H_0$: $c_{TISAC} = c_{other}$.

Finally, we evaluated the suitability for BOLD analysis. Table 3.5 shows the comparison of the normalized cross-correlation between estimated CDFs of the phase values before and after applying SACs. In practice, it is desirable that SAC methods maintain the BOLD responses. The results indicate that the BOLD responses of all three correction methods are not different from those of uncorrected data (one-way ANOVA for 3T dataset: $F_{2,969} = 1.03$, $p$-value = 0.358; one-way ANOVA for 7T dataset: $F_{2,78} = 0.34$, $p$-value = 0.710).

3.4.6 Discussion

The experimental results indicate that SAC methods can correct geometric distortions in EPIs, even when these distortions are severe as in the 7T dataset. HySCO produces corrected images with ghost artifacts around the brain boundary. TOPUP produces images with good distortion corrections and no ghost
3.5. Chapter summary

artifacts, but it introduces blur and affects the BOLD responses. Judged by visual inspection and the performance measures, TISAC produces output images with better alignment to the structural image, compared to TOPUP and HySCO, especially for the dataset with severe geometric distortions.

For high spatial-resolution fMRI, the blurring effects on post-processing are of great concern. We found that all the SAC methods add blur into the corrected images, but the proposed TISAC method adds the least amount. Furthermore, the blur that SAC methods add to the motion-corrected data is much less than the blur that the motion correction step adds to the original data. Evaluation of the structural similarity indicates that the mutual information measure is able to reflect obvious improvements between uncorrected and corrected images.

3.5 Chapter summary

This chapter introduced a novel method, called TISAC, to correct the susceptibility artifacts in high spatial resolution EPI-fMRI images. The proposed method uses a reversed-PE EPI-fMRI image pair and a $T_{1w}$ image. The symmetric registration principle is adopted to combine the reversed-PE images and produce the corrected images that aligns well with the $T_{1w}$ image. The $T_{1w}$ image is used to regularize the registration and select the hyper-parameters via Bayesian optimization.

Evaluating TISAC on two high spatial-resolution EPI-fMRI datasets shows that TISAC outperforms the existing methods in terms of accuracy and robustness, particularly in sub-millimeter images obtained by the high field scanner. The proposed method produces sharper corrected images with better geometric correction. It is effective in preserving the structure of the $T_{1w}$ image in regions of significant SA distortions. Furthermore, the proposed method requires less computational resources than TOPUP and HySCO methods. The corrected images produced by TISAC provide better results in subsequent fMRI analysis, while still keeping the BOLD responses as found in the uncorrected images.
Chapter 4

Susceptibility Artifact Correction with Unsupervised Deep Learning

Chapter contents

4.1 Introduction ................................................. 48
4.2 Related work .................................................. 49
4.3 Unsupervised deep learning framework for SAC .......... 51
  4.3.1 CNN architecture for mapping ......................... 53
  4.3.2 Spatial transformation unit for unwarping .......... 54
  4.3.3 Loss function for self-supervised network learning ... 55
4.4 Experiments and analysis .................................... 57
  4.4.1 Brain EPI datasets .................................... 57
  4.4.2 Experimental methods ................................ 59
  4.4.3 Analysis of hyper-parameters ......................... 60
  4.4.4 Comparison with other methods in correction accuracy .. 61
  4.4.5 Comparison with other methods in processing speed ... 65
  4.4.6 Discussion .............................................. 67
4.5 Chapter summary ............................................. 68
4.1 Introduction

The existing SAC methods are time-consuming as they implement an iterative-optimization algorithm. These SAC methods, thus, are unsuitable for time-sensitive applications, for example correction on an MRI scanner. To reduce the processing time, this chapter\textsuperscript{a} introduces an unsupervised deep learning technique, called S-Net, for correcting the susceptibility artifacts in 3D reversed-PE images. A convolutional neural network (CNN) is used to map a pair of reversed-PE images to the displacement field in the PE direction. Note that for the rest of this chapter, the term displacement field represents the displacement in the PE direction for every voxel. Then, a differentiable spatial transform unit is used to unwarp the input (distorted) image pair via the predicted displacement field. S-Net is trained in an end-to-end manner using a training set of reversed-PE image pairs. After training, correcting a new 3D reversed-PE image pair is achieved by simply evaluating the trained S-Net on the given input images. This approach, therefore, avoids the highly computational cost of the existing iterative-optimization approaches.

The contributions of this chapter are highlighted as follows:

1. We design a convolutional encoder-decoder network to map a 3D reversed-PE image pair to the displacement field. The network consists of an encoder for image downsampling and a decoder for image upsampling. The encoder uses a series of convolutional (conv) layers and leaky rectified linear units (LeakyReLUs) to extract hierarchical image contents. The decoder uses a series of conv layers, LeakyReLUs, and upsampling layers to recover the full-resolution image features and estimate the displacement field. A spatial transform unit is designed to unwarp the 3D input images along the PE direction and produce an output image pair. To the best of our knowledge,\textsuperscript{a}

\textsuperscript{a}Chapter 4 has been published in: “An unsupervised deep learning technique for susceptibility artifact correction in reversed phase-encoding EPI images,” Magnetic Resonance Imaging, vol. 71, pp. 1-10, 2020.
the proposed technique is the first attempt at using a learning-based approach with an convolutional encoder-decoder to correct the susceptibility artifacts.

2. We explore an unsupervised learning strategy in design the proposed S-Net. The term unsupervised learning arises from the fact that S-Net is trained without additional ground-truth information, e.g. the “desired” displacement field or the “desired” corrected images, which are impractical to acquire. In our approach, S-Net is trained by maximizing the similarity of the output image pairs and the smoothness of the displacement field, which are inspired by the traditional reversed-PE based SAC methods.

3. We evaluate the performance of the proposed technique and compare it with existing SAC methods using three datasets. The datasets include one dataset acquired by our team using a 7T scanner, and two public datasets acquired using a 3T scanner and published by the Human Connectome Project [47]. The experimental results show that our unsupervised S-Net provides the corrected images which are comparable to results of state-of-the-art SAC methods while requiring fewer computational resources and no additional data, such as structural images.

The remainder of this chapter is organized as follows. Section 4.2 presents a literature review on learning-based image registration methods. Section 4.3 introduces the proposed self-supervised learning framework. Section 4.4 presents experiments and analysis of the proposed method and the related methods. Finally, Section 4.5 summarizes this chapter.

4.2 Related work

Recall that the reserved-PE SAC first estimates the displacement field based on a pair of images acquired using an identical sequence but with opposite PE di-
4.2. Related work

The corrected images are then obtained by unwarping the distorted images via the estimated displacement field. Since the reversed-PE SAC can be cast as a symmetric non-rigid image registration problem, this section presents an overview of the learning-based non-rigid registration.

Non-rigid image registration is typically formulated as an optimization problem to seek a non-linear correspondence for every pixel (or voxel) between a moving image and a fixed image. This approach could be slow due to the required iterative optimization. Learning-based registration is an approach to avoid the iterative optimization. The recent development in convolutional neural networks has shown remarkable successes in solving the image registration problem [72, 73, 74].

Several learning-based image registration methods have been proposed. Those methods usually consist of two parts: (i) a network to map the input image pair (fixed image and moving image) to displacement fields\(^b\); and (ii) a spatial transform unit to unwarp the input moving image via the predicted displacement fields. In term of ground-truth requirement for training, we group the registration methods into two categories: supervised registration and unsupervised registration.

The supervised registration methods usually train the mapping network by minimizing the difference between the output and desired displacement fields [75, 76, 77, 78, 79]. The desired displacement fields can be real [78]; however, most of them are synthesized using conventional registration methods, such as LDDMM shooting algorithm [76, 80] and SimpleITK\(^c\) [77]. The mapping network can be designed as a patch-based CNN [75, 76, 77], a dual-stream CNN [78], or a convolutional encoder-decoder [79]. The supervised registration methods present a promising direction. However, artificial displacement fields are tedious to acquire and can restrict the types of deformations.

\(^b\)The image registration estimates the displacements in all dimensions for every pixel (voxel). In the 3D image registration, thus, three displacement fields are estimated.

\(^c\)http://www.simpleitk.org/
4.3 Unsupervised deep learning framework for SAC

The unsupervised registration methods often combine the mapping network and the spatial transform unit into an end-to-end model, which takes an input image pair (fixed image and moving image) and produces a correction of the moving image. The network is trained with the loss function reflecting the smoothness of the displacement fields and the similarity between the input fixed image and the corrected moving image [81, 82, 83, 84]. The mapping network can be designed as a fully convolutional network [81, 82], or a convolutional encoder-decoder [83, 84]. The spatial transform unit can be implemented with deconvolutional operators [81], a cubic spline interpolation scheme [82], or a linear interpolation scheme [83, 84]. Note that the methods with the interpolation schemes are inspired by the spatial transformer network [85].

4.3 Unsupervised deep learning framework for SAC

This chapter introduces a deep learning technique, called S-Net, for correcting a pair of 3D reversed-PE images. Fig. 4.1 illustrates an overview of the proposed S-Net. It consists of two parts: (i) a mapping function to estimate the displacement field $U$ from a pair of 3D reversed-PE images $I_1$ and $I_2$; and (ii) a spatial transform unit to recover the corrected images by unwarping the input images with the estimated 3D displacement field. Note that the computations in S-Net are performed in 3D.

The mapping function $f_{\theta}: (I_1, I_2) \rightarrow U$ is modeled by a deep CNN, where $\theta$ is the network parameters. Here, $U$ is the 3D displacement field along the PE direction of image $I_1$ to the corrected image. Due to the inverse distortion property in the PE direction of the SAs, the displacement field of image $I_2$ to the corrected image is $-U$. The detail of the CNN architecture is presented in Section 4.3.1.

The spatial transform unit unwarpes images $I_1$ and $I_2$ using the computed displacement field. The unwarping returns the corrected images, expressed as $[I_1 \otimes (G + U)]$ and $[I_2 \otimes (G - U)]$, where $G$ is the identity transform (i.e. the regular
4.3. Unsupervised deep learning framework for SAC

Figure 4.1: The proposed learning framework (S-Net) for correcting the SAs in reversed-PE images. S-Net accepts a pair of reversed-PE images and produces the displacement field and the corrected images.

In the training phase, the network parameters $\theta$ are learned to produce the best estimation of the displacement field and the corrected images. The displacement field is good if unwarping the input images using it produces the corrected images as similar to each other as possible. The goal of training is to minimize a loss function with respect to the network parameters $\theta$ on a training dataset. In this work, the loss function is unsupervised. It captures the similarity of the corrected images and the smoothness of the displacement field. The detail of the unsupervised loss function is presented in Section 4.3.3.

Although the proposed S-Net is inspired by the VoxelMorph network presented in [83], we provide several new contributions. First, the loss function in [83] is computed from the input image and the output image of VoxelMorph, while our loss function is computed from the two output images of S-Net. Second, VoxelMorph is designed for registering a moving $T_{1w}$ image to a fixed $T_{1w}$ image, while S-Net is designed for correcting the susceptibility artifacts in a pair of EPI images acquired with reversed phase encoding. In VoxelMorph, the deformation is caused by affine transformations or anatomy differences between subjects. In
4.3. Unsupervised deep learning framework for SAC

contrast, in S-Net the deformation is caused by intrinsic magnetic susceptibility differences of tissues when putting in a magnetic fields.

4.3.1 CNN architecture for mapping

The mapping CNN architecture used in our method (see Fig. 4.2) is inspired by U-Net [86], VoxelMorph [83], and DL-GP [87]. It accepts a two-channel image formed by concatenating two 3D reversed-PE images, and produces a 3D displacement field of the same size as the input images. The mapping network consists of an encoder (left side) and a decoder (right side). Both the encoder and decoder use a kernel size of 3 for their 3D conv layers.

The encoder consists of four units E1, E2, E3 and E4; each unit consists of a conv layer and a LeakyReLU with a slope coefficient of 0.2. In the encoder, the conv layers have a stride of 2 to reduce the size of their input feature maps by half. With this scheme, succeeding units of the encoder extract hierarchical features of the input image pair.

The decoder consists of seven units D1 to D7. Units D1, D2, D3, and D4 each consists of a conv layer, a LeakyReLU, and an upsampling layer. Each upsampling layer doubles the size of its feature maps, so the output feature maps of D4 have the same size as the input images $I_1$ and $I_2$. Units D5, D6, D7 each consists of a
conv layer and a LeakyReLU. To retain more local information when upsampling and reduce the effects of vanishing gradients in training, skip connections are introduced in the network, as indicated in Fig. 4.2. For example, the 3D input images $I_1$ and $I_2$ are concatenated to the outputs of D5 before being fed to the next unit. The output feature maps of E1 are concatenated to the outputs of D3 before being fed to the next unit.

### 4.3.2 Spatial transformation unit for unwarping

This section presents the spatial transformation unit (STU), which is used to unwarp the distorted images and form the corrected images. S-Net is trained by minimizing the loss function via a gradient-based algorithm, so the STU must be differentiable. Inspired by the spatial transformer network [85], we construct a differential STU unit, as illustrated in Fig. 4.3.

![Figure 4.3: Spatial transformation unit for unwarping distorted image via the displacement field.](image)

In unwarping operations, the value for voxel $p = (p_x, p_y, p_z)$ of the corrected image $[I \otimes (G + U)]$ is taken from the distorted image $I$ at voxel $p' = (p_x + U(p), p_y, p_z)$. This chapter uses the convention that the PE direction is in the first dimension. Because $p'$ may have continuous coordinates, image interpolation is required. In this chapter, we use the linear interpolation to estimate the output voxel value:

$$[I \otimes (G + U)](p) = [1 - (p'_x - q_{1,x})]I(q_1) + [1 - (q_{2,x} - p'_x)]I(q_2), \quad (4.1)$$
4.3. Unsupervised deep learning framework for SAC

where \( q_1 \) and \( q_2 \) are two neighbors of voxel \( p' \) along the first dimension with \( q_{1,x} \leq p'_{x} < q_{2,x} \). The linear interpolation is selected due to its differentiability and computational efficiency.

### 4.3.3 Loss function for self-supervised network learning

This section introduces an unsupervised loss function to train the S-Net model. Note that the proposed loss function is based on the outputs of S-Net, i.e. the displacement field and two output images. It consists of two components: \( L_{\text{sim}} \) and \( L_{\text{smooth}} \). \( L_{\text{sim}} \) penalizes the differences between two estimated-corrected images, whereas \( L_{\text{smooth}} \) penalizes the local spatial variations in \( U \). For 3D input images \( I_1 \) and \( I_2 \), let \( E_1 \) and \( E_2 \) be the two corresponding 3D output (corrected) images. The loss function is defined as

\[
L(I_1, I_2, U) = L_{\text{sim}}(E_1, E_2) + \lambda L_{\text{smooth}}(U), \tag{4.2}
\]

where \( \lambda \) is a positive regularization parameter.

In this chapter, we use the local cross-correlation (LCC) for \( L_{\text{sim}} \) for two reasons. First, LCC is suitable to measure the similarity of images with the same modality (i.e. the corrected images). Second, LCC is more robust to intensity variations found across scans and datasets [88].

The LCC measure can be explained as follows. Consider an image \( X \). Let \( \bar{X} \) be the local mean image obtained by applying an \( n \times n \times n \) averaging filter on \( X \). The local mean-removed image \( \hat{X} \) is computed as

\[
\hat{X} = X - \bar{X}. \tag{4.3}
\]

For a given voxel \( p \), let \( W(p) \) denote the set of voxels in the \( n \times n \times n \) volume centered on \( p \). For a pair of images \( E_1 \) and \( E_2 \), we compute an image \( C \):
4.3. Unsupervised deep learning framework for SAC

\[ C(p) = \frac{\left( \sum_{p_i \in W(p)} \hat{E}_1(p_i) \hat{E}_2(p_i) \right)^2}{\sum_{p_i \in W(p)} [\hat{E}_1(p_i)]^2 \sum_{p_i \in W(p)} [\hat{E}_2(p_i)]^2}. \]  \hspace{1cm} (4.4)

The LCC measure for 3D images \( E_1 \) and \( E_2 \) is now computed as

\[ \text{LCC}(E_1, E_2) = \sum_p C(p), \] \hspace{1cm} (4.5)

where the summation is over all image voxels. A higher LCC indicates more similarity between two output images.

We now can express \( \mathcal{L}_{\text{sim}}(E_1, E_2) \) as

\[ \mathcal{L}_{\text{sim}}(E_1, E_2) = 1 - \text{LCC}(E_1, E_2). \] \hspace{1cm} (4.6)

Minimizing \( \mathcal{L}_{\text{sim}} \) increases the similarity between the output images \( E_1 \) and \( E_2 \), but it may generate a non-smooth and physically unrealistic displacement field \( U \). In this chapter, we enforce the smoothness of the displacement field \( U \) using the diffusion regularizer on the spatial gradient of the displacement field \( U \)

\[ \mathcal{L}_{\text{smooth}}(U) = \sum_{p \in \Omega} ||\nabla U(p)||^2, \] \hspace{1cm} (4.7)

where \( \nabla U(p) = \left( \frac{\partial U(p)}{\partial x}, \frac{\partial U(p)}{\partial y}, \frac{\partial U(p)}{\partial z} \right) \). The voxel difference is used to approximate \( \nabla U(p) \), for example

\[ \frac{\partial U(p)}{\partial x} \approx U(p_x + 1, p_y, p_z) - U(p_x, p_y, p_z). \] \hspace{1cm} (4.8)

The proposed S-Net is trained by optimizing the loss function with respect to (w.r.t.) the trainable parameters \( \theta \). This is often done using stochastic optimization algorithms. In this chapter, we adopt the adaptive moment estimation, known as Adam, [89]. This algorithm starts with an initial guess of \( \theta \). The next estimate of
4.4 Experiments and analysis

\( \theta \) is computed iteratively as

\[
\theta_{t+1} = \theta_t - \alpha \frac{m_t}{\sqrt{v_t} + \epsilon},
\]

(4.9)

where subscript \( t \) is the iteration number, \( \alpha \) is a positive learning rate, \( m_t \) is the estimator of the first bias-corrected moment, and \( v_t \) is the estimator of the second bias-corrected moment. Here, \( \epsilon \) is a very small positive scalar used to avoid the division by zero resulted from the vanishing gradients. Let \( \nabla_\theta L_t \) be the gradient of the loss function on the mini-batch at step \( t \) w.r.t. the parameters \( \theta \). The estimators at step \( t \) are defined as

\[
m_t = \beta_1 m_{t-1} + \frac{(1 - \beta_1) \nabla_\theta L_t}{1 - (\beta_1)^t},
\]

(4.10)

\[
v_t = \beta_2 v_{t-1} + \frac{(1 - \beta_2) \left[ \nabla_\theta L_t \right]^2}{1 - (\beta_2)^t},
\]

(4.11)

where hyper-parameters \( \beta_1 \) and \( \beta_2 \) are the exponential decay rates for the first and the second estimators. Kingma and Ba suggested the following hyper-parameter values: \( \beta_1 = 0.9 \) and \( \beta_2 = 0.999 \) [89].

4.4 Experiments and analysis

This section presents the experiments and analysis of the proposed S-Net. Section 4.4.1 describes the datasets used in the experiments, and Section 4.4.2 explains the experimental methods. Section 4.4.3 analyzes the regularization and learning rate parameters. Sections 4.4.4 and 4.4.5 compare the correction accuracy and processing time of the proposed method with other representative methods, respectively. Finally, Section 4.4.6 discusses the experimental results.

4.4.1 Brain EPI datasets

The S-Net framework was evaluated using three datasets: fMRI-7T, fMRI-3T, and DWI-3T. The datasets are diverse in the acquisition sequence, modality, distortion
4.4. Experiments and analysis

Table 4.1: A summary of the datasets used in the experiments.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>No. subj.</th>
<th>Gender dist.</th>
<th>Age dist.</th>
<th>Image size</th>
<th>Resolution</th>
<th>Acq. sequences</th>
<th>Field strength</th>
<th>PE directions</th>
</tr>
</thead>
<tbody>
<tr>
<td>fMRI-7T</td>
<td>3</td>
<td>Males: 3</td>
<td>26-30 years: 1&lt;br&gt;31-35 years: 1&lt;br&gt;over 36 years: 1</td>
<td>192 × 192 × 48</td>
<td>0.833 × 0.833 × 0.810 mm³</td>
<td>3D-GRE EPI (WIP1080 [60])</td>
<td>7T</td>
<td>LR and RL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Females: 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fMRI-3T</td>
<td>182</td>
<td>Males: 72</td>
<td>22-25 years: 24&lt;br&gt;26-30 years: 85&lt;br&gt;31-35 years: 71&lt;br&gt;over 36 years: 2</td>
<td>90 × 104 × 72</td>
<td>2 × 2 × 2 mm³</td>
<td>Single-band 2D spin-echo EPI</td>
<td>3T</td>
<td>LR and RL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Females: 110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DWI-3T</td>
<td>180</td>
<td>Males: 71</td>
<td>22-25 years: 23&lt;br&gt;26-30 years: 84&lt;br&gt;31-35 years: 71&lt;br&gt;over 36 years: 2</td>
<td>144 × 168 × 111</td>
<td>1.25 × 1.25 × 1.25 mm³</td>
<td>Single-band 2D spin-echo EPI</td>
<td>3T</td>
<td>LR and RL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Females: 109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: LR = left-to-right; RL = right-to-left.

property, field strength, resolution, image size, and dataset size. A summary of these datasets is presented in Table 4.1.

The fMRI-7T is a sub-millimeter dataset, which was acquired by our team for a retinotopic mapping fMRI study. It includes data of three healthy subjects acquired using a 7T scanner. Each subject was scanned in seven to ten experimental runs with the PE direction from left-to-right. Each subject was also scanned in two short runs with the PE direction from right-to-left, 20 s each. Each experimental run produced 183 or 187 images, and each short run produced 10 images. All images were motion-corrected using SPM12 tools [65]. To remove redundancy, we computed the mean image of each scan. The mean image of each experimental run was paired with the mean image of its respective inverse, resulting in 25 reversed-PE image pairs in total.

The fMRI-3T is a subset of the unpreprocessed 3T fMRI data in the public HCP dataset for studying functional connectivity of the human brain [48]. It includes data of 182 healthy subjects from the group Subjects with 7T MR Session Data. Each subject was scanned in 11 to 18 runs, using a customized Siemens 3T scanner. Each run contained two spin-echo images acquired by an identical sequence but with the PE direction alternated between RL and LR.

The DWI-3T is a subset of the unpreprocessed 3T DWI data in the public HCP dataset for reconstructing the complex axonal fiber architecture [48, 90]. It includes
4.4. Experiments and analysis

data of 180 healthy subjects from the group Subjects with 7T MR Session Data. Each subject has three pairs of reversed-PE images. Each image pair was acquired using an identical spin-echo sequence but with the PE direction alternated between RL and LR. The fMRI-3T and DWI-3T datasets used different spin-echo sequences, and consequently had different image modalities.

4.4.2 Experimental methods

This section describes the S-Net implementation and the experimental setup. To implement S-Net, we used Keras [91] with TensorFlow as the backend [92].

To evaluate the S-Net models, we split each dataset into a training set and a test set, as summarized in Table 4.2. The training set was used to select the hyper-parameters and train S-Net models, and the test set was used to evaluate the network performance. The experiments were conducted on image pairs of the datasets directly, without any pre-processing step. Furthermore, the reversed phase-encoding images are acquired by the MRI scanners, and they follow several physical constraints. Thus, we did not use any data augmentation in training the TS-Net model.

Table 4.2: A summary of the training set and test set for each of the three datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Training set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. subjects</td>
<td>No. pairs</td>
</tr>
<tr>
<td>fMRI-7T</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>fMRI-3T</td>
<td>156</td>
<td>2774</td>
</tr>
<tr>
<td>DWI-3T</td>
<td>150</td>
<td>436</td>
</tr>
</tbody>
</table>

To analyze the hyper-parameters, including the learning rate $\alpha$ of the Adam optimizer and the regularization parameter $\lambda$ of the loss function in (4.2), we trained the S-Net model with multiple trials of hyper-parameter values. The trials were selected using the tree of parzen estimator (TPE) algorithm from the Hyperopt library [58, 93, 94]. The objective function used in the TPE algorithm was the validation loss, which is the mean of the LCC measures between the pairs of output images from subsets of the training sets. For each dataset, the
4.4. Experiments and analysis

hyper-parameter values with the best validation loss were selected for training S-Net. Network training was done with a batch size of 3 for 1500 epochs.

To measure the correction accuracy of the proposed method, we computed the corrected images of the test sets using S-Net and two state-of-the-art SAC methods, i.e. TOPUP [34] and TISAC [14]. The S-Net output images were then compared with the uncorrected images, and the output images produced by the two existing SAC methods. For accuracy comparison, we considered two types of similarity: (i) EPI similarity between the pair of reversed-PE images; and (ii) anatomical similarity between the EPI image and its corresponding $T_{1w}$ image. The normalized mutual information (NMI) is used to measure the similarity. The NMI value range is from 0 (no mutual information) to 1 (complete correlation). In summary, we used two accuracy measures: EPI-NMI and anatomical-NMI.

To evaluate the processing speed, for each of the three SAC methods, we measured the times for two main tasks: (i) correcting the distorted image pair; and (ii) unwarping the distorted images given the displacement field. For S-Net, image correction was done by evaluating the trained model with the input reversed-PE images, and unwarping was done via the spatial transformation unit. This experiment was conducted on the DWI-3T dataset. As shown in Table 4.1, this dataset has the largest image size and also the second largest number of subjects. All timings were reported on an Ubuntu 18.04.3 LTS workstation with an Intel Core i5-9600K CPU at 3.6 GHz, 32 GB of RAM, and an NVIDIA GeForce RTX2080 GPU with 8 GB memory.

4.4.3 Analysis of hyper-parameters

This section analyzes the effects of two hyper-parameters of the model: the learning rate $\alpha$ of the Adam optimizer and the regularization parameter $\lambda$ of the objective function in (4.2). Fig. 4.4 shows the validation loss for different values of $\alpha$ and $\lambda$ for the three datasets. The dot represents the values of the pair $(\alpha, \lambda)$ produced by the TPE algorithm. The color of the dot encodes the validation loss:
a dark color denotes a low validation loss, and a light color denotes a high validation loss. In each plot, the star (*) indicates the value pair \((a, \lambda)\) that provides the lowest validation loss for the dataset.

**Figure 4.4**: Validation loss of S-Nets versus the regularization parameter \(\lambda\) and learning rate \(\alpha\). The value pairs \((a, \lambda)\) are found by the TPE algorithm. The color indicates the value of the validation loss; the star indicates the pair \((a, \lambda)\) with the lowest validation loss.

Fig. 4.4 shows that for the fMRI-7T dataset, the validation loss is small when \(a \in (0.0003, 0.001)\) and \(\lambda \in (0, 0.0001)\). For the fMRI-3T dataset, the validation loss is small when \(a \in (0, 0.0001)\) and \(\lambda \in (0, 2)\). For the DWI-3T dataset, the validation loss is small when \(a \in (0, 0.0001)\) and \(\lambda \in (0, 1)\). These results show that there are several choices for \((a, \lambda)\) that can give acceptable performances. The suitable values for \((a, \lambda)\) as indicated by the TPE algorithm are summarized in Table 4.3.

**Table 4.3**: Hyper-parameter values used in training the S-Net models for the three datasets.

<table>
<thead>
<tr>
<th>Hyper-params</th>
<th>fMRI-7T</th>
<th>fMRI-3T</th>
<th>DWI-3T</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>5.6e-4</td>
<td>7.87e-5</td>
<td>8e-5</td>
</tr>
<tr>
<td>(\lambda)</td>
<td>0.017</td>
<td>1.344</td>
<td>0.333</td>
</tr>
</tbody>
</table>

### 4.4.4 Comparison with other methods in correction accuracy

This section presents both visual results and quantitative metrics of the proposed S-Net technique on the test sets. Figs. 4.5 and 4.6 shows one representative slice of uncorrected and corrected images from each of the three test sets. Each example
4.4. Experiments and analysis

includes the two reversed-PE images (Rows 1 and 2) and the absolute difference of
the image pair (Row 3). It can be seen from Figs. 4.5 and 4.6 that S-Net removes the
distortions in the uncorrected images. In addition, S-Net produces the comparable
output images as TOPUP and TISAC. Note that, the SAC methods work with 3D
images; however, for simplicity, 2D slices are presented in the examples.

Figure 4.5: Representative visual results of SAC methods from the fMRI-7T dataset.
Column 1: input uncorrected images. Columns 2, 3, and 4: output corrected images
produced by TOPUP, TISAC, and S-Net, respectively. Rows 1 and 2: left-right and right-
left phase-encoding EPI images, respectively. Row 3: absolute difference of LR and RL
images.

Tables 4.4 and 4.5 summarize the statistical comparison of S-Net versus the
uncorrected data (i.e. no correction) and the two SAC methods on the test sets. The
p-values are computed by the paired t-tests on the NMI measures between the S-
Net outputs and the compared images. The null hypothesis is \( H_0 : m_{S-Net} = m_{other} \).
A p-value smaller than 0.05 indicates that the null hypothesis is rejected at a
confidence level of 95%. In other words, there is a significant difference in the
Figure 4.6: Representative visual results of SAC methods from the HCP 3T datasets.
4.4. Experiments and analysis

NMI measures between the S-Net outputs and the compared images.

(a) NMI of EPI to EPI

(b) NMI of EPI to T₁w

Figure 4.7: Comparisons of the accuracy between the S-Net technique versus other two existing SAC methods on the test sets. **Left column**: similarity between the corrected EPI images. **Right column**: similarity between the T₁w image and corrected EPI images. Because of differences in the datasets, the plots are drawn in different y-axis ranges for clarity.

Statistical results in Tables 4.4 and 4.5 indicate that the proposed S-Net improves both the EPI-NMI and anatomical-NMI measures of the uncorrected images significantly. S-Net improves the NMI measures in most cases when compared to the other SAC methods. However, differences in the mean NMI measures between S-Net and the two SAC methods are small (see the visual comparisons
4.4. Experiments and analysis

### Table 4.4: Statistical comparisons of the EPI-NMI measures between the proposed S-Net versus the uncorrected data and other two existing SAC methods on the test sets.

<table>
<thead>
<tr>
<th>Datatypes</th>
<th>fMRI-7T</th>
<th>fMRI-3T</th>
<th>DWI-3T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean ± std</td>
<td>p-value</td>
<td>mean ± std</td>
</tr>
<tr>
<td>Uncorrected</td>
<td>0.302 ± 0.003</td>
<td>0.000</td>
<td>0.442 ± 0.023</td>
</tr>
<tr>
<td>TOPUP</td>
<td>0.943 ± 0.008</td>
<td>0.000</td>
<td>0.995 ± 0.001</td>
</tr>
<tr>
<td>TISAC</td>
<td>0.990 ± 0.001</td>
<td>0.000</td>
<td>0.992 ± 0.002</td>
</tr>
<tr>
<td>S-Net</td>
<td>0.976 ± 0.002</td>
<td>0.994</td>
<td>0.992 ± 0.000</td>
</tr>
</tbody>
</table>

The null hypothesis $H_0: m_{S-Net} = m_{other}$. A $p$-value below $0.05$ indicates that the null hypothesis is rejected at a confidence level of $95\%$.

### Table 4.5: Statistical comparisons of the anatomical-NMI measures between the proposed S-Net versus the uncorrected data and other two existing SAC methods on the test sets.

<table>
<thead>
<tr>
<th>Datatypes</th>
<th>fMRI-7T</th>
<th>fMRI-3T</th>
<th>DWI-3T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean ± std</td>
<td>p-value</td>
<td>mean ± std</td>
</tr>
<tr>
<td>Uncor.</td>
<td>0.147 ± 0.014</td>
<td>0.000</td>
<td>0.733 ± 0.014</td>
</tr>
<tr>
<td>TOPUP</td>
<td>0.259 ± 0.002</td>
<td>0.000</td>
<td>0.917 ± 0.011</td>
</tr>
<tr>
<td>TISAC</td>
<td>0.250 ± 0.001</td>
<td>0.000</td>
<td>0.920 ± 0.013</td>
</tr>
<tr>
<td>S-Net</td>
<td>0.254 ± 0.003</td>
<td>0.924</td>
<td>0.924 ± 0.012</td>
</tr>
</tbody>
</table>

The null hypothesis $H_0: m_{S-Net} = m_{other}$. A $p$-value below $0.05$ indicates that the null hypothesis is rejected at a confidence level of $95\%$.

in Fig. 4.7). In summary, S-Net is able to remove the geometric distortions in the reversed-PE images, and produce the corrected images at the same accuracy level of TOPUP and TISAC.

### 4.4.5 Comparison with other methods in processing speed

This section compares the speed of the proposed S-Net method versus the two state-of-the-art SAC methods. Table 4.6 presents the average processing time of SAC methods for correcting a pair of reversed-PE images. To correct one reversed-PE image pair, S-Net required an average of 2.80 s, whereas TOPUP required an average of 1033.03 s and TISAC required an average of 56.49 s. Hence, S-Net is 369 times faster than TOPUP and 20 times than TISAC. Furthermore, when using the GPU, S-Net took only only 0.96 s to produce the corrected images. To the best of our knowledge, there is no GPU implementation for TOPUP and TISAC methods.

Table 4.7 shows the average runtimes taken to unwarp a pair of distorted images. To unwarp an image pair, TOPUP and TISAC required about 3.50 s, while S-Net on CPU required an average of 1.75 s, i.e. about two times faster.
4.4. Experiments and analysis

**Table 4.6:** Elapsed time (in second) of SAC methods for correcting a pair of reversed-PE images with a size of $144 \times 168 \times 111$.

<table>
<thead>
<tr>
<th>Processor</th>
<th>TOPUP (mean ± std.)</th>
<th>TISAC (mean ± std.)</th>
<th>S-Net (mean ± std.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPU</td>
<td>-</td>
<td>-</td>
<td>0.96 ± 0.07</td>
</tr>
<tr>
<td>CPU</td>
<td>1033.03 ± 382.03</td>
<td>56.49 ± 11.71</td>
<td>2.80 ± 0.10</td>
</tr>
</tbody>
</table>

than TOPUP and TISAC. S-Net on GPU required an average of 1.33 s, *i.e.* about 2.5 times faster than TOPUP and TISAC.

**Table 4.7:** Elapsed time (in second) of SAC methods for unwarping separately two reversed-PE images with a size of $144 \times 168 \times 111$.

<table>
<thead>
<tr>
<th>Processor</th>
<th>TOPUP (mean ± std.)</th>
<th>TISAC (mean ± std.)</th>
<th>S-Net (mean ± std.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPU</td>
<td>-</td>
<td>-</td>
<td>1.33 ± 0.03</td>
</tr>
<tr>
<td>CPU</td>
<td>3.65 ± 0.06</td>
<td>3.40 ± 0.06</td>
<td>1.75 ± 0.09</td>
</tr>
</tbody>
</table>

The processing time results demonstrate that S-Net is orders of magnitude faster than both TOPUP and TISAC for correcting a reserved-PE image pair. Unwarping by the spatial transform unit employed by S-Net require less time than the unwarping procedures used in TOPUP and TISAC. A reason for the fast processing is that once trained, the proposed S-Net can directly correct the reversed-PE images. In contrast, existing methods, such as TOPUP and TISAC, require an additional optimization step for each test pair of images. Furthermore, the proposed S-Net has a GPU implementation which can run significantly faster than the CPU implementation.

For completeness, Table 4.8 summarizes the training time of the proposed S-Net model for the three datasets. Note that this training time is incurred only once, and it does not affect the time for correcting images in the test phase. In contrast, as shown in Table 4.6 the existing methods require significant processing time (iterative optimization) for each new test image.
4.4. Experiments and analysis

Table 4.8: Training time (in second) of the proposed S-Net model on different datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Image size</th>
<th>1 epoch (mean ± std.)</th>
<th>1500 epochs</th>
</tr>
</thead>
<tbody>
<tr>
<td>fMRI-7T</td>
<td>192 × 192 × 48</td>
<td>18.58 ± 0.24</td>
<td>27,874.11 (0.32 days)</td>
</tr>
<tr>
<td>fmri-3T</td>
<td>90 × 104 × 72</td>
<td>238.90 ± 2.41</td>
<td>358,356.89 (4.15 days)</td>
</tr>
<tr>
<td>DWI-3T</td>
<td>144 × 168 × 111</td>
<td>291.87 ± 4.97</td>
<td>437,804.00 (5.07 days)</td>
</tr>
</tbody>
</table>

4.4.6 Discussion

This section discusses the proposed S-Net in three aspects: feasibility, scalability, and limitations. In terms of the feasibility, the proposed S-Net is able to learn the features of the reversed-PE images from the training set. The trained S-Net model can then be applied to unseen data to obtain the corrected images. The experiment results on the three different datasets show that S-Net is not limited by the image resolution, image size, image modality, and training set size. Furthermore, it can cope with different acquisition sequences, such as spin-echo, gradient-echo, sequences optimized for fMRI (i.e. short TE, no diffusion gradient), and sequences optimized for DWI (i.e. long TE, diffusion gradients).

In terms of the scalability, the trained S-Net can produce equally good corrected images, in comparison to the state-of-the-art SAC methods. Importantly, to correct a pair of distorted images, S-Net takes only 3 s using CPU or 1 s using GPU. This fast processing is desirable as it opens new applications. For example, the trained S-Net models can be integrated into the MRI scanner to correct SAs in near real-time, while the traditional reversed-PE SAC methods are too slow for this purpose. The trained S-Net models can be used to correct large-scale datasets, such as the HCP with 1200 subjects, in a reasonable time and on modest computing platforms. For example, to correct 1200 image pairs with a size of 144 × 168 × 111 voxels, TOPUP requires an average of 1,239,636 s (about 14 days), while S-Net needs only 1,152 s (about 19 minutes) on GPU, or 3,360 s (56 minutes) on CPU.
4.5. Chapter summary

The proposed S-Net still has some limitations. First, as with deep learning it requires significant time to train, especially for large-scale training sets. Second, arbitrary input images need to be resized to the image size selected for training the S-Net model. However, robustness to image sizes can be enhanced by training S-Net on resized images from various sources. Furthermore, the training time does not affect the time for correcting the test image pairs.

4.5 Chapter summary

This chapter introduced a novel unsupervised deep learning technique, S-Net, to correct susceptibility artifacts in the reversed-PE EPI image pairs in an end-to-end setting. The proposed S-Net contains a convolutional encoder-decoder to map a reversed-PE image pair to the displacement field. The displacement field is then fed to spatial transform units to unwarp the input images, resulting in the corrected images. S-Net is trained in an unsupervised manner, without requiring the ground-truth data. The loss function of S-Net is derived from its outputs to reflect the similarity of the corrected images, and the smoothness of the displacement field.

Evaluations on three different datasets demonstrate that the proposed S-Net can provide corrected images that are comparable with those provided by the state-of-the-art SAC methods, i.e. TOPUP and TISAC. Notably, it runs significantly faster than the traditional SAC methods: about 369 times faster than TOPUP and 20 times faster than TISAC. This speed improvement allows new applications of the proposed S-Net, such as the integration into the MRI scanner console.
Chapter 5

Susceptibility Artifact Correction with Multi-modal Deep Learning

Chapter contents

5.1 Introduction .......................................................... 69
5.2 Deep learning framework for SAC ................................. 71
  5.2.1 Deep network for mapping ................................ 72
  5.2.2 Loss function for TS-Net training ......................... 73
5.3 Experiments and analysis .......................................... 76
  5.3.1 Image datasets and experimental methods ............. 76
  5.3.2 Ablation study of the proposed method ............... 78
  5.3.3 Comparison with other methods ......................... 82
5.4 Chapter summary ..................................................... 87

5.1 Introduction

This chapter presents a multi-modal deep learning method, called TS-Net, to correct the susceptibility artifacts in a pair of 3D reversed-PE images. TS-Net contains a convolutional encoder-decoder to map a 3D reversed phase-encoding

\footnote{Chapter 5 has been submitted to *IEEE Transactions on Biomedical Imaging*, 2020.}
image pair to a 3D displacement field, and a spatial transform unit to unwarp the input (distorted) image pair based on the predicted displacement field. The proposed TS-Net extends our previous SAC framework presented in Chapter 4. The new contributions of this chapter are highlighted as follows:

1. We design a convolutional encoder-decoder to map a 3D reversed-PE image pair to a 3D displacement field. In the encoder and decoder, batch normalization (BN) layers are used to reduce changes in the distribution of the convolutional layers’ inputs. We also introduce customized layers to make TS-Net work with different image sizes while maintaining the spatial resolution of the input images. Thus, the proposed convolutional encoder-decoder is robust to different image sizes, resolutions, and modalities (including 3T and 7T EPI).

2. We extend the estimate of the displacement field to all three dimensions instead of only along the phase-encoding direction. In other words, TS-Net predicts the displacement field that captures the 3D displacements for every voxel. This, to our knowledge, is a significant improvement to most existing SAC methods [14, 34, 95], which estimate the distortions only along the PE direction and ignore the distortions along other directions.

3. We introduce a learning method that leverages T1-weighted (T1w) images in training TS-Net; the term multi-modal in the name of the proposed method arises from the fact that it uses both EPI and T1w modalities. The motivation is that the T1w image is widely considered as a gold standard representation of a subject’s brain anatomy [50], and it is readily available in brain studies [7]. To make TS-Net more applicable for general use, the T1w image is used only in training for network regularization, but not in the inference phase.

4. We provide an extensive evaluation of the proposed TS-Net on four large public datasets from the Human Connectome Project [47]. First, an ablation study is conducted to analyze the effects of using different similarity
measures to train TS-Net, the effects of various components of the TS-Net framework, and the effects of using a pre-trained TS-Net when training for a new dataset. Second, TS-Net is compared with three state-of-the-art SAC methods, i.e. TOPUP [34], TISAC Chapter 3, and S-Net Chapter 4, in terms of correction accuracy and processing time.

The remainder of this chapter is organized as follows. Section 5.2 introduces the proposed TS-Net method. Section 5.3 presents the experiments and analysis of the proposed method in comparison with other existing methods. Finally, Section 5.4 summarizes this chapter.

5.2 Deep learning framework for SAC

This chapter introduces a multi-modal deep learning framework, called TS-Net, to correct the susceptibility artifacts in a pair of 3D reversed-PE images. Fig. 5.1 illustrates an overview of the proposed TS-Net. It consists of two parts: (i) a mapping function to predict the displacement field $U$ from a pair of 3D reversed-PE images $I_1$ and $I_2$; and (ii) a 3D spatial transform unit to obtain the corrected images by unwarping the input images with the predicted displacement field. In contrast to Chapters 3 and 4, here we estimate a 3D displacement field, or three displacement values for each voxel. Thus, the displacement field $U$ can be represented as $[U_x, U_y, U_z]$, where $U_d$ is the displacement field in the $d$ direction.

The mapping function $f_\theta : (I_1, I_2) \rightarrow U$ is modeled by a convolutional encoder-decoder, where $\theta$ is the set of network parameters. Here, $U$ is the displacement field of image $I_1$ to the corrected image. Due to the inverse distortion property, the displacement field of image $I_2$ to the corrected image is $-U$. The architecture of the convolutional encoder-decoder is presented in Section 5.2.1.

The 3D spatial transform unit, which is inspired by the VoxelMorph work [84], unwarp images $I_1$ and $I_2$ using the computed displacement field. The unwarping returns the corrected images, expressed as $[I_1 \otimes (G + U)]$ and $[I_2 \otimes (G - U)]$, where
5.2. Deep learning framework for SAC

Figure 5.1: The proposed learning framework (TS-Net) for correcting the SAs in reversed-PE images. The TS-Net accepts a pair of 3D reversed-PE images and produces the displacement fields and the corrected images.

\[ G = [G_x, G_y, G_z] \] is the identity transforms \( i.e. \) the regular grids in the \( x, y, \) and \( z \) directions), and \( \otimes \) is the interpolation operator. In this chapter, we use linear interpolation due to its differentiability and computational efficiency.

In the training phase, the network parameters \( \theta \) are learned to produce a good estimation of the displacement field and two accurate corrected images. The goal of training is to minimize a loss function with respect to the network parameters \( \theta \) on a training dataset. In this chapter, the loss function captures the similarity between the output images, the smoothness of the displacement field, and the similarity between the \( T_{1w} \) and the output images. The loss function is presented in Section 5.2.2.

5.2.1 Deep network for mapping

In this chapter, a deep neural network is used to map the input image pair to the displacement field. The deep network used in our method (see Fig. 5.2) is inspired by convolutional encoder-decoder Chapter 4, U-Net [86], and DL-GP [87]. It accepts a two-channel image formed by concatenating two 3D reversed-PE
images, and produces a displacement field of the same size as the input images. The mapping network consists of an encoder (left side) and a decoder (right side).

Figure 5.2: The convolutional encoder-decoder for mapping a pair of reversed-PE images to the displacement fields. Box: output feature maps of a layer. Number inside each box: the number of feature maps in the layer. Number below each box: feature map size relative to the full input image size.

To make TS-Net cope with different input image sizes, we add a size-normalization layer before the encoder and a size-recovery layer after the decoder. The size-normalization layer uses zero-padding so that each input dimension is divisible by 16. The size-recovery layer crops the decoder output to the size of the input image. To resize images, TS-Net uses zero-padding instead of interpolation because the displacements in the EPI images are small and sensitive to image interpolation.

The encoder-decoder is inherited from the S-Net architecture in Chapter 4. To mitigate changes in the distribution of the convolutional layer’s input, each convolutional layer is followed by a BN layer [96].

5.2.2 Loss function for TS-Net training

The loss function used for training the proposed TS-Net consists of three terms. The first term, $L_{\text{sim}}$, promotes the similarity between the pair of corrected images. The second term, $L_{\text{smooth}}$, enforces the local smoothness of the predicted displace-
ment field. The third term, \( L_{\text{anat}} \), enhances the structural alignment between the corrected EPI images and the \( T_{1w} \) image.

For input reversed-PE images \( I_1 \) and \( I_2 \), and a \( T_{1w} \) image \( A \), let \( E_1 \) and \( E_2 \) be the two corresponding output (corrected) images. The loss function is defined as

\[
L(I_1, I_2, A, U) = L_{\text{sim}}(E_1, E_2) + \lambda L_{\text{smooth}}(U) + \gamma L_{\text{anat}}(E_1, E_2, A),
\]

(5.1)

where \( \lambda \) and \( \gamma \) are regularization parameters. These positive and user-defined regularization parameters represent the trade-off between the similarity of the corrected images, the smoothness of the displacement field, and the similarity of the \( T_{1w} \) image to the output images.

In this chapter, we investigate three possible similarity measures for unimodal images: mean squared error (MSE), local cross-correlation (LCC) [88], and local normalized cross-correlation (LNCC) [97].

- The **MSE** between two images \( E_1 \) and \( E_2 \) is defined as

\[
\text{MSE}(E_1, E_2) = \frac{1}{|\Omega|} \sum_{p \in \Omega} [E_1(p) - E(p)]^2,
\]

(5.2)

where \( \Omega \) is the image domain. A smaller value of MSE indicates a higher similarity between the images. Thus, the \( L_{\text{sim}} \) loss based on the MSE measure is

\[
L_{\text{sim}}^{\text{MSE}}(E_1, E_2) = \text{MSE}(E_1, E_2).
\]

(5.3)

- The **LCC** metric, which is described in Section 4.3.3, is proportional to the similarity. Thus, the \( L_{\text{sim}} \) loss based on the LCC measure is defined as

\[
L_{\text{sim}}^{\text{LCC}}(E_1, E_2) = 1 - \text{LCC}(E_1, E_2).
\]

(5.4)
5.2. Deep learning framework for SAC

- The LNCC can be defined as follows. Let $\tilde{X}$ be the variance image of $X$:

$$
\tilde{X}(p) = \sum_{p_i \in W(p)} [X(p_i)]^2 - \frac{1}{n^3} \left[ \sum_{p_i \in W(p)} X(p_i) \right]^2.
$$

(5.5)

Let $R$ be the correlation image between two images $E_1$ and $E_2$:

$$
R(p) = \sum_{p_i \in W(p)} [E_1(p_i)E_2(p_i)] - \frac{1}{n^3} \sum_{p_i \in W(p)} E_1(p_i) \sum_{p_i \in W(p)} E_2(p_i).
$$

(5.6)

The LNCC between two images $E_1$ and $E_2$ is given by

$$
\text{LNCC}(E_1, E_2) = \frac{1}{|\Omega|} \sum_{p \in \Omega} \frac{[R(p)]^2}{\tilde{E}_1(p) \tilde{E}_2(p)}.
$$

(5.7)

A higher LNCC indicates higher similarity between two output images. We now can express the $L_{\text{sim}}$ loss based on the LNCC measure as

$$
L^{\text{LNCC}}_{\text{sim}}(E_1, E_2) = 1 - \text{LNCC}(E_1, E_2).
$$

(5.8)

Minimizing $L_{\text{sim}}$ increases the similarity between the output images $E_1$ and $E_2$, but it may generate a non-smooth and physically unrealistic displacement field $U$. In this chapter, we enforce the smoothness of a displacement field $U$ using a diffusion regularizer on the spatial gradients of the displacement field:

$$
L_{\text{smooth}}(U) = \sum_{p \in \Omega} ||\nabla U(p)||^2,
$$

(5.9)

where $\nabla U(p) = \left( \frac{\partial U(p)}{\partial x}, \frac{\partial U(p)}{\partial y}, \frac{\partial U(p)}{\partial z} \right)$. The voxel difference is used to approximate $\nabla U(p)$, for example

$$
\frac{\partial U(p)}{\partial x} \approx U(p_x + 1, p_y, p_z) - U(p_x, p_y, p_z).
$$

(5.10)

To improve the training phase, a $T_{1w}$-based regularizer, which penalizes the distances from the corrected images to the corresponding $T_{1w}$ structural image, is
5.3. Experiments and analysis

added to the loss function. The normalized mutual information (NMI) is used to measure the similarity between the output images and the $T_{1w}$ image because it is effective for multi-modal images. The range of NMI value is from 0 (no mutual information) to 1 (complete correlation). Thus, the loss function for the structural similarity is defined as

$$L_{anat}(E_1, E_2, A) = 1 - \frac{NMI(E_1, A) + NMI(E_1^c, A)}{2}.$$  \hfill (5.11)

The proposed TS-Net is trained by optimizing the loss function with respect to (w.r.t.) the trainable parameters $\theta$. In this chapter, we use Adam [89] algorithm to find the optimal trainable parameters $\theta$. Please see the detail of Adam in Section 4.3.3.

5.3 Experiments and analysis

This section presents the experiments and analysis of the proposed TS-Net. Section 5.3.1 describes the brain image datasets and experimental methods. Section 5.3.2 presents an ablation study of TS-Net. Finally, Section 5.3.3 compares the proposed method with other representative SAC methods in terms of correction accuracy and processing time.

5.3.1 Image datasets and experimental methods

The TS-Net framework was evaluated using four datasets: fMRI-3T, DWI-3T, fMRI-7T, and DWI-7T. These datasets were extracted from the public Human Connectome Project repository in the group Subjects with 7T MR Session Data [48]. The evaluation data consist of two functional MRI datasets for studying functional connectivity of the human brain [48], and two diffusion MRI datasets for reconstructing the complex axonal fiber architecture [48, 90]. The four datasets are diverse in the acquisition sequence, imaging modality, distortion property, field strength, resolution, image size, and dataset size. A summary of the four...
5.3. Experiments and analysis

datasets is presented in Table 5.1. Note that the apparent diffusion coefficient (ADC) map was not acquired in the DWI datasets. The b-values were 1000, 2000, and 3000 s/mm$^2$ for the DWI-3T dataset, and 1000 and 2000 s/mm$^2$ for the DWI-7T dataset.

<table>
<thead>
<tr>
<th>Table 5.1: A summary of the datasets used in the experiments.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Datasets</strong></td>
</tr>
<tr>
<td>fMRI-3T</td>
</tr>
<tr>
<td>DWI-3T</td>
</tr>
<tr>
<td>fMRI-7T</td>
</tr>
<tr>
<td>DWI-7T</td>
</tr>
</tbody>
</table>

Abbreviations: LR = left-to-right; RL = right-to-left; AP = anterior-to-posterior; PA = posterior-anterior.

To implement TS-Net, we used Keras [91] with TensorFlow as the backend [92]. For training TS-Net, the Adam optimizer was used with the learning rate $\alpha = 0.001$, and the exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$, as suggested by Kingma and Ba [89]. The regularization parameters $\lambda$ and $\gamma$ were selected using the Tree of Parzen Estimator algorithm from the Hyperopt library [58, 93, 94]. The batch sizes and regularized parameters used in training TS-Net are shown in Table 5.2. The batch sizes and regularization parameters used in training TS-Net are shown in Table 3. We used a batch size of 1 due to memory limitation for three datasets: DWI-3T, fMRI-7T, and DWI-7T. We used a batch size of 4 for the fMRI-3T dataset; in this case each batch was formed as four image pairs selected randomly from the training set.

To evaluate TS-Net, for each dataset we first split the subjects randomly into two parts: A and B. Then, the training set was formed by selecting randomly reversed-PE image pairs of each subject in Part A; this strategy reduces the data
Table 5.2: Values of hyper-parameters in training TS-Net on the four datasets.

<table>
<thead>
<tr>
<th>Params</th>
<th>fMRI-3T</th>
<th>DWI-3T</th>
<th>fMRI-7T</th>
<th>DWI-7T</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>0.1771</td>
<td>0.002</td>
<td>0.9323</td>
<td>0.025</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Batch size</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

repetition of subjects. The test set was formed from all reversed-PE pairs of each subject in Part B. Table 5.3 summarizes the training sets and test sets for the four datasets. The training sets were used to select the hyper-parameters and train the TS-Net models, and the test sets were used to evaluate the correction accuracy of the TS-Net models.

Table 5.3: A summary of the training set and test set for each of the four datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Training set</th>
<th>Test set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. subjects</td>
<td>No. pairs</td>
</tr>
<tr>
<td>fMRI-3T</td>
<td>156</td>
<td>1872</td>
</tr>
<tr>
<td>DWI-3T</td>
<td>150</td>
<td>436</td>
</tr>
<tr>
<td>fMRI-7T</td>
<td>153</td>
<td>3212</td>
</tr>
<tr>
<td>DWI-7T</td>
<td>148</td>
<td>155</td>
</tr>
</tbody>
</table>

The experiments were conducted on image pairs of the datasets directly, without any pre-processing step. The experiments for evaluating processing times were performed on a system that has an Intel Core i5-9600K CPU at 3.6 GHz, 32 GB of RAM, and an NVIDIA GeForce RTX2080 GPU with 8 GB memory. The other experiments were performed on a system that has an Intel Xero Gold 5115 CPU at 2.4 GHz, and an NVIDIA GeForce GTX Titan Xp with 12 GB memory.

5.3.2 Ablation study of the proposed method

This section analyzes the proposed TS-Net method on five aspects: (i) effects of using different similarity measures; (ii) effects of the network configurations in TS-Net; (iii) effects of using the 3D distortion model and \( T_{1w} \) guidance; (iv) effects of using a pre-trained TS-Net in training other datasets; and (v) the visualization of the predicted displacement field.
5.3. Experiments and analysis

Effects of similarity measures in network training: In this experiment, for each training set, we trained three TS-Net models each with a similarity loss: (i) MSE; (ii) LCC; and (iii) LNCC. The effects of using different similarity measures were evaluated in two aspects: the validation loss and the training time of each epoch. The validation loss was measured as the mean similarity measures for output image pairs across subsets of the training sets. We conducted the experiments on the four datasets: fMRI-3T, DWI-3T, fMRI-7T, and DWI-7T.

Fig. 5.3 shows the validation loss versus time when training TS-Net with the similarity loss as MSE, LCC, and LNCC. It can be seen that TS-Net trained with the LNCC measure produces the lowest validation loss, while TS-Net trained with the MSE measure produces the highest validation loss. TS-Nets trained with the LNCC and LCC measures produce a competitive LCC validation loss on two datasets (DWI-3T and fMRI-7T). Considering the validation loss versus the training time, it is clear that the LNCC measure is a better choice than the MSE and the LCC for training TS-Net. Based on this experiment, the LNCC measure was subsequently used as the similarity loss for all the remaining experiments.

![Figure 5.3](image-url)

*Figure 5.3: Validation loss of the models trained with three types of similarity loss (MSE, LCC, and LNCC) versus training time (in second) on the four datasets: (a) fMRI-3T; (b) DWI-3T; (c) fMRI-7T; and (d) DWI-7T. Top row: validation loss in terms of MSE. Middle row: validation loss in terms of LCC. Bottom row: validation loss in terms of LNCC.*
5.3. Experiments and analysis

Effects of the network configurations in TS-Net: In this experiment, we analyzed the effects of four different network configurations: (i) TS-Net without batch normalization and with upsampling layer (UL) (ii) TS-Net with instance normalization (IN) [98], and with UL; (iii) TS-Net with BN and transposed convolution (TC) [99]; and (iv) TS-Net with BN and UL (proposed method). The validation loss during the training phase was computed as the average LNCC measure between the output image pairs, across subsets of the training sets. This validation loss was then used to compare different network configurations.

Fig. 5.4 shows the validation loss versus the training time on three datasets: fMRI-3T, DWI-3T, and DWI-7T; each subfigure includes the validation loss for the four network configurations. Several observations can be made. First, using batch normalization (proposed TS-Net, green curve) provides a lower validation loss compared to not using batch normalization (blue curve). Second, using batch normalization (proposed TS-Net, green curve) provides a similar or lower validation loss compared to using instance normalization (orange curve). Third, using the upsampling layer (proposed TS-Net, green curve) has a similar validation loss compared to using the transpose convolution (magenta curve). These results justify our selected configuration for TS-Net.

![Figure 5.4: Validation loss of four trained models versus training time (in second) on three datasets: (a) fMRI-3T; (b) DWI-3T; and (c) DWI-7T. The four models are: (i) TS-Net without batch normalization and with upsampling layer (UL); (ii) TS-Net with instance normalization and UL; (iii) TS-Net with batch normalization (BN) and transposed convolution; and (iv) TS-Net with BN and UL (proposed method).](image)

Effects of using the 3D distortion model and anatomical guidance by $T_{1w}$: In this experiment, we trained three types of networks: (i) TS-Net with the 1D distortion
model as used in S-Net Chapter 4; (ii) TS-Net with 3D distortion model and without $T_{1w}$ guidance; and (iii) TS-Net with the 3D distortion model and $T_{1w}$ guidance (proposed method). Fig. 5.5 shows the validation loss versus the training time on three datasets: fMRI-3T, DWI-3T, and DWI-7T. Several observations can be made. First, the proposed TS-Net with $T_{1w}$ guidance (green-solid curve) has lower validation losses than the TS-Net without $T_{1w}$ guidance (brown dash-dotted curve). This result shows that incorporating $T_{1w}$ guidance can improve the correction accuracy. Second, the proposed TS-Net using the 3D distortion model (green-solid curve) produces significantly lower validation losses than TS-Net using the 1D distortion model (magenta-dashed curve). This result shows that the 3D distortion model used in the proposed TS-Net provides more accurate correction than the 1D distortion model (i.e. only along the phase-encoding direction), which is used in S-Net and existing iterative-optimization SAC methods.

Effects of using a pre-trained TS-Net: In this experiment, we explored whether using a TS-Net model pre-trained on one dataset can reduce the training time on another dataset, compared to a randomly initialized TS-Net. To this end, we trained two TS-Net models: (i) from scratch; and (ii) using an initial network, which had been pre-trained for 1500 epochs on the fMRI-3T dataset.

Fig. 5.6 shows the validation loss versus training time on three datasets: DWI-3T, fMRI-7T, and DWI-7T. The figure shows that the validation loss when training
TS-Net using a pre-trained model (cyan dash-dotted curve) is much lower than when training from scratch (green-solid curve). The result suggests that TS-Net is able to learn generalized features for correcting the susceptibility artifacts from one dataset. Subsequently, these learned features could be adopted to reduce the validation loss on other datasets. In other words, the training time on a new dataset, which is a common bottleneck of deep learning systems, can be reduced by using a pre-trained TS-Net as the network initialization.

Figure 5.6: Validation loss of two trained TS-Net models versus training time (in second) on three datasets: (a) DWI-3T; (b) fMRI-7T; and (c) DWI-7T. The two TS-Net models were trained: (i) from scratch; and (ii) using the pre-trained model of the fMRI-3T dataset.

**Visualization of the predicted displacement fields:** Fig. 5.7 shows the samples of the displacement field estimated by the trained TS-Net for the four test sets. The displacement field is shown in three directions (left-right, anterior-posterior, and superior-interior). The visual results indicate that TS-Net is able to predict realistic displacement fields. Furthermore, TS-Net can estimate the geometric distortions along the directions that are not the dominant PE direction.

### 5.3.3 Comparison with other methods

This section compares the correction accuracy and processing speed of TS-Net with three state-of-the-art SAC methods, i.e. TOPUP [34], TISAC Chapter 3, and S-Net Chapter 4. To evaluate the correction accuracy of the proposed method, we trained S-Net and TS-Net for a maximum of 1500 epochs. The trained models were used to compute the corrected image pairs of the test sets. For TOPUP and TISAC, the corrected image pairs were obtained by implementing the iterative
5.3. Experiments and analysis

(a) fMRI-3T
(b) DWI-3T
(c) fMRI-7T
(d) DWI-7T

Figure 5.7: Samples of three predicted displacement fields (in voxel) of TS-Net from the four test sets. In each subfigure, *left image*: displacement field in the left-right (LR) direction; *middle image*: displacement field in the anterior-posterior (AP) direction; and *right image* displacement field in the superior-interior (SI) direction. The dominant phase-encoding dimension (direction) is shown in red text; the other two other dimensions are shown in white text.

optimization. Here, the correction accuracy is measured as the LNCC similarity between the pair of reversed-PE images.

Fig. 5.8 shows sample slices of uncorrected and corrected images from each of the four test sets. Each example includes two reversed-PE images (Rows 1 and 2) and the absolute difference between the two images (Row 3). The arrows indicate the regions where TS-Net produces significantly improved correction in comparison with other three SAC methods. It can be seen that TS-Net removes distortions in the uncorrected images significantly. In general, TS-Net produces the output images that are comparable to or better than the outputs of TOPUP, TISAC, and S-Net. Note that the SAC methods work with 3D images; however, for visualization, 2D slices are presented in the figures.

Table 5.4 summarizes the statistical accuracy comparison of TS-Net outputs versus the uncorrected images, and TS-Net outputs versus outputs of the three SAC methods on the test sets. The *p*-values were computed by paired t-tests on the LNCC measures between TS-Net outputs and each of four image types: uncorrected images, TOPUP outputs, TISAC outputs, and S-Net outputs. The
null hypothesis is $H_0 : m_{\text{S-Net}} = m_{\text{other}}$. All $p$-values are smaller than 0.001, which indicates that the null hypothesis is rejected at a confidence level of 99.9%. In other words, the TS-Net output images have significant differences (i.e. improvements).

Figure 5.8: Sample visual results of SAC methods from the four test sets. In each subfigure, Column 1: input uncorrected images. Columns 2, 3, 4, and 5: output corrected images produced by TOPUP, TISAC, S-Net, and TS-Net, respectively. Rows 1 and 2: reversed phase-encoding EPI images. Row 3: color bar of the absolutely different maps. Row 4: absolute difference between the image pair. Row 5: corresponding T$_1$w image of the reversed-PE images and the estimated displacement fields of the compared SAC methods. For visualization, only the displacement field in the phase-encoding direction of TS-Net is shown. Row 6: color bar of the displacement fields, in which the number expresses the number of voxels shifted.
5.3. Experiments and analysis

compared to the output images of other methods.

For visual clarity, Fig. 5.9 shows the box plots for comparing the LNCC measures of the four SAC methods. The results in Table 5.4 and Fig. 5.9 show three notable trends. First, TS-Net produces output images that have significantly higher LNCC measures than the uncorrected images; in other words, TS-Net does reduce the susceptibility artifacts. Second, TS-Net produces output images that have higher LNCC measures than the outputs of the TISAC method in 4 out of 4 datasets, and the outputs of the TOPUP methods in 3 out of 4 datasets. This means that TS-Net has better correction accuracy compared to the two iterative-optimization methods, TISAC and TOPUP. Third, TS-Net also produces higher LNCC measures than S-Net in 4 out of 4 datasets. Compared to S-Net, the proposed TS-Net has several differences, one of which is its use of $T_{1w}$ images in training. This result demonstrates that including the gold-standard representation of a subject’s brain anatomy helps regularize the susceptibility artifact correction in TS-Net. Note that TS-Net does not require the $T_{1w}$ image in the inference phase, which explains its comparable processing speed with S-Net, as analyzed next.

**Table 5.4:** Statistical comparisons on the test sets of the LNCC measures between the TS-Net outputs versus four image types: uncorrected images, TOPUP outputs, TISAC outputs, and S-Net outputs.

<table>
<thead>
<tr>
<th>Datatypes</th>
<th>fMRI-3T</th>
<th>DWI-3T</th>
<th>fMRI-7T</th>
<th>DWI-7T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean ± std</td>
<td>p-value</td>
<td>mean ± std</td>
<td>p-value</td>
</tr>
<tr>
<td>Uncorrected</td>
<td>0.335 ± 0.023</td>
<td>0.000</td>
<td>0.142 ± 0.020</td>
<td>0.000</td>
</tr>
<tr>
<td>TOPUP</td>
<td>0.753 ± 0.024</td>
<td>0.000</td>
<td>0.468 ± 0.031</td>
<td>0.000</td>
</tr>
<tr>
<td>TISAC</td>
<td>0.674 ± 0.036</td>
<td>0.000</td>
<td>0.436 ± 0.058</td>
<td>0.000</td>
</tr>
<tr>
<td>S-Net</td>
<td>0.608 ± 0.027</td>
<td>0.000</td>
<td>0.242 ± 0.039</td>
<td>0.000</td>
</tr>
<tr>
<td>TS-Net</td>
<td>0.692 ± 0.022</td>
<td>0.000</td>
<td>0.571 ± 0.034</td>
<td>0.000</td>
</tr>
</tbody>
</table>

The null hypothesis $H_0: m_{TS-Net} = m_{other}$. A p-value below 0.05 indicates that the null hypothesis is rejected at a confidence level of 95%; in other words, the similarity measure (LNCC) of TS-Net is significant different than of the compared method.

To compare the processing speed, we first randomly selected 50 distorted image pairs for each of the four datasets. We then recorded the time for correcting the selected image pairs by four SAC methods: TOPUP, TISAC, S-Net, and TS-Net. Table 5.5 shows the average processing time per an image pair of TS-Net and the three SAC methods. Over the four datasets, TS-Net is 396.72 times faster than TOPUP, 29.45 times faster than TISAC, and only 1.05 times slower than S-Net.
5.3. Experiments and analysis

![Box plots of LNCC-based accuracy for different datasets](image)

**Figure 5.9:** Comparisons of the proposed TS-Net versus other three SAC methods in terms of the LNCC-based accuracy on the test sets. Due to differences in LNCC ranges of the datasets, the plots are drawn in different y-axis ranges for clarity. In each box plot, the top line is the maximum LNCC value excluding the outliers; the bottom line is the minimum LNCC value excluding the outliers; the middle line is the median LNCC value; the solid rectangle is the interquartile range of the LNCC values; and the points are the outliers.

Both deep learning-based SAC methods (TS-Net and S-Net) can be accelerated by five times using the GPU instead of the CPU. Note that in the experiments for all datasets, the proposed TS-Net has 260,187 trainable parameters whereas the S-Net model has 259,241 trainable parameters. In other words, the proposed TS-Net requires only 0.36% more trainable parameters than S-Net.

The results of TS-Net over the four datasets show that the inference time of TS-Net is linearly proportional to the size of the input images. To correct an image pair with a size of $90 \times 104 \times 72$, TS-Net takes 0.65 s using CPU, and 0.14 s using GPU. On average, the inference speed of TS-Net is approximately 1.08 million voxels per second with CPU, and 5.98 million voxels per second with GPU.
Table 5.5: Processing time (in second) of SAC methods for correcting a pair of reversed-PE images.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Processor</th>
<th>fMRI-3T (mean ± std)</th>
<th>DWI-3T (mean ± std)</th>
<th>IMRI-7T (mean ± std)</th>
<th>DWI-7T (mean ± std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOPUP</td>
<td>CPU</td>
<td>90 × 104 × 72</td>
<td>144 × 168 × 111</td>
<td>130 × 130 × 85</td>
<td>200 × 200 × 132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>252.55 ± 3.61</td>
<td>997.39 ± 9.04</td>
<td>535.71 ± 44.29</td>
<td>1944.65 ± 18.72</td>
</tr>
<tr>
<td>TISAC</td>
<td>CPU</td>
<td>25.76 ± 11.81</td>
<td>57.73 ± 12.03</td>
<td>28.48 ± 5.14</td>
<td>126.13 ± 26.25</td>
</tr>
<tr>
<td>S-Net</td>
<td>CPU</td>
<td>0.63 ± 0.03</td>
<td>2.21 ± 0.03</td>
<td>1.36 ± 0.03</td>
<td>4.55 ± 0.04</td>
</tr>
<tr>
<td>TS-Net</td>
<td>CPU</td>
<td>0.65 ± 0.04</td>
<td>2.30 ± 0.05</td>
<td>1.45 ± 0.04</td>
<td>4.92 ± 0.06</td>
</tr>
<tr>
<td>S-Net</td>
<td>GPU</td>
<td>0.13 ± 0.14</td>
<td>0.42 ± 0.18</td>
<td>0.22 ± 0.16</td>
<td>0.72 ± 0.25</td>
</tr>
<tr>
<td>TS-Net</td>
<td>GPU</td>
<td>0.14 ± 0.16</td>
<td>0.43 ± 0.21</td>
<td>0.23 ± 0.18</td>
<td>0.80 ± 0.26</td>
</tr>
</tbody>
</table>

For completeness, Table 5.6 summarizes the time to train the proposed TS-Net from scratch on the four datasets. Note that the training time is incurred only once, and it does not affect the time for correcting image pairs.

Table 5.6: Training time (in second) of the proposed TS-Net model on different datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>No. pairs</th>
<th>1 epoch (mean ± std.)</th>
<th>1500 epochs</th>
</tr>
</thead>
<tbody>
<tr>
<td>fMRI-3T</td>
<td>1872</td>
<td>535.39 ± 14.23</td>
<td>803,089.86 (9.30 days)</td>
</tr>
<tr>
<td>DWI-3T</td>
<td>436</td>
<td>466.03 ± 20.86</td>
<td>699,047.41 (8.09 days)</td>
</tr>
<tr>
<td>fMRI-7T</td>
<td>3212</td>
<td>1808.36 ± 287.43</td>
<td>2,712,540.54 (31.40 days)</td>
</tr>
<tr>
<td>DWI-7T</td>
<td>155</td>
<td>389.62 ± 21.27</td>
<td>584,429.00 (6.76 days)</td>
</tr>
</tbody>
</table>

5.4 Chapter summary

This chapter presented a high-speed multi-modal deep learning framework, TS-Net, to correct the susceptibility artifacts in reversed phase-encoding 3D EPI image pairs in an end-to-end setting. The proposed TS-Net contains a convolutional encoder-decoder to map the reversed-PE image pair to the 3D displacement field. The predicted displacement field is then fed into a spatial transform unit to unwarp the input images, resulting in the corrected 3D EPI images. In the training phase, the proposed TS-Net utilizes $T_{1w}$ images to regularize the susceptibility artifact correction. However, the $T_{1w}$ image is not used in the inference phase to simplify the use of TS-Net.

Evaluation on the four large datasets demonstrates that the proposed TS-Net
provides higher correction accuracy than TISAC and S-Net in all four datasets, and than TOPUP in three out of four datasets. Over the four datasets, TS-Net runs significantly faster than the iterative-optimization SAC methods: 396.72 times faster than TOPUP and 29.45 times faster than TISAC. TS-Net is slightly slower than S-Net, but it still meets the real-time correction requirement of MRI scanners. Furthermore, the training time of TS-Net on a new dataset can be reduced by using a pre-trained model.
Susceptibility artifact correction is an important step in the image processing pipeline of the brain study using the EPI images. Some existing SAC methods require the modification of the MRI scanner hardware and acquisition protocol. Other SAC methods are time-consuming due to the required iterative-optimization. In this thesis, we propose three methods to correct the susceptibility artifacts in the reversed-PE image pair. Our proposed methods do not require the modification of the scanner hardware or acquisition protocol. The proposed methods reduce the processing time for SAC significantly. It is a step towards correcting the susceptibility artifacts directly inside the MRI scanners via a real-time implementation.

This chapter is organized as follows: Section 6.1 summarizes the research contributions of the thesis; Section 6.2 outlines the future work and research directions; Section 6.3 draws conclusion for the thesis.
6.1 Research summary

The research activities have been documented in several chapters of the thesis. They are summarized as follows.

- We provided the literature review on the principles of MRI and the related work for susceptibility artifact corrections. We reviewed the strength and weakness of the existing SAC methods. [Chapter 2].

- We proposed a new iterative-optimization algorithm, named TISAC, using the $T_1w$ regularization for correcting the susceptibility artifacts in a reversed-PE image pair. The $T_{1w}$ image is used to enforce the structural validity of the corrected images, and to select the regularization parameters via Bayesian optimization. The experimental results show that TISAC outperforms state-of-the-art methods in terms of accuracy and robustness, particularly in sub-millimeter fMRI-EPI images obtained by the high-field scanner. The proposed TISAC produces sharper corrected images with better geometric correction. It is effective in preserving the structure of the $T_{1w}$ image in regions of significant distortions. Notably, TISAC requires significantly fewer computational resources than state-of-the-art methods. Furthermore, the TISAC corrected images provide better results in subsequent functional brain analysis. [Chapter 3].

- We developed a novel unsupervised deep learning framework, named S-Net, for correcting the susceptibility artifacts in reversed-PE image pair. The proposed S-Net contains a convolutional encoder-decoder to map a reversed-PE image pair to the displacement field. The displacement field is then fed into spatial transform units to unwarp the input images, resulting in the corrected images. S-Net is trained in an unsupervised manner, without requiring the ground-truth data. The loss function of S-Net is derived from its outputs to reflect the similarity of the corrected images, and the smooth-
ness of the displacement field. The experimental results demonstrate that the proposed S-Net technique can provide corrected images that are comparable with those provided by the state-of-the-art SAC methods. Notably, it runs significantly faster than the iterative-optimization SAC methods: about 369 times faster than TOPUP and 20 times faster than TISAC. [Chapter 4].

- We proposed a multi-modal deep learning framework, named TS-Net, to obtain a higher correction accuracy. The proposed method extends S-Net in three aspects: (i) the encoder-decoder is designed to cope with different input image sizes and be robust with the variant shift of the feature maps; (ii) the distortions are estimated in all directions rather than only in the PE direction; and (iii) the structural information of the $T_{1w}$ image is used as a regularizer in training the improved network. Since the improved network uses two modalities, EPI and $T_{1w}$, in training, it is called the multi-modal deep learning framework. The experimental results indicate that the proposed TS-Net can provide corrected images with higher correction accuracy compared to TISAC and S-Net. TS-Net provides a higher correction accuracy than TOPUP on almost all evaluated datasets. Furthermore, the training time of TS-Net on a dataset is reduced significantly by initializing it as a pre-trained network from another dataset. This indicates that the proposed TS-Net is able to utilize transfer learning from different datasets. [Chapter 5].

6.2 Future work

Possible research directions can be summarized as follows:

- Develop a supervised deep learning framework using Gaussian process regressor (GPR) to correct the susceptibility artifacts in a pair of reversed-PE image. In recent years, the convolutional neural network integrated with a GPR has attracted the attention of researchers due to its promising inference
results. Therefore, extending of this thesis to supervised deep learning framework with a GPR would be invaluable.

- Explore a neural architecture search algorithm to find an optimum convolutional network for mapping the reversed-PE image pair to the displacement field. The optimum network is measured in terms of three factors: (i) SAC correction accuracy; (ii) network size; and (iii) inference time. The main advantage of neural architecture search algorithms is that they are able to automate the network design process and produce novel, high-performing network structures.

- Investigate a transfer learning algorithm for SAC deep networks. Currently, the two proposed S-Net and TS-Net networks are trained from scratch, which requires significant time especially for large-scale training sets. Because a good initialization for weights can reduce the training time significantly; it is worth to develop a transfer learning algorithm to optimize the training process.

6.3 Conclusion

This thesis has presented the methods for correcting susceptibility artifacts in reversed-PE image pair. The proposed SAC methods include: (i) an iterative-optimization algorithm (TISAC) using the $T_{1w}$ regularizer; (ii) an unsupervised deep learning framework (S-Net); and (iii) a multi-modal deep learning framework (TS-Net). First, the experimental results show that the proposed TISAC removes the susceptibility artifacts with the same correction accuracy as the existing SAC methods while maintaining the sharpness of the input images and performing much faster. Second, the proposed deep learning S-Net produces the corrected images with the same accuracy as the state-of-the-art SAC methods while running many times faster. Third, the proposed TS-Net produces the corrected images with higher accuracy than the state-of-the-art SAC methods.
6.3. Conclusion

TS-Net is more robust than S-Net; it can work with different modalities, MRI resolutions, and input image sizes. The training time of TS-Net can be reduced significantly by initializing the trainable weights with a network pre-trained on a different dataset. Notably, S-Net and TS-Net methods reduce the time to correct a distorted image pair significantly because they produce the output images directly and non-iteratively, by evaluating the trained network. Moreover, S-Net and TS-Net can utilize the GPU implementation to reduce further the processing time. The processing time results of S-Net and TS-Net indicate that it is possible to integrate S-Net and TS-Net into MRI scanners that require the real-time implementation.
Magnetic resonance imaging is an invasive technique for producing images of inside the biological samples. All MRI techniques rely on a physical phenomenon called Nuclear Magnetic Resonance to generate the MR signals, and the image formation technique to generate the image of the object being scanned. In this Appendix, the principles of MR signal are first presented in Section A.1. The principles of MR image formation are then presented in Section A.2.

A.1 MR signal generation

The magnetic resonance signal generation relies on the magnetic property of living tissues and the NMR phenomenon, which was discovered and developed by Rabi, Purcell et al., and Bloch [100, 101, 102, 103, 104]. The NMR phenomenon is the emission of the radio-frequency radiation when placing a nuclear in a strong magnetic field and then applying a radio-frequency pulse.

Note that the NMR phenomenon is applied for the nuclear possessed the NMR property which has an odd number of protons, neutrons, or both. The nuclear with the NMR property has an intrinsic property called spin, which the NMR is based on. For this reason, the nuclear with the NMR property is called a
A.1. MR signal generation

spin \([15, 105]\). In natural, there is a few nuclei with the NMR property, e.g. \(^1\text{H}, ^{13}\text{C}, ^{14}\text{N}, ^{23}\text{Na}, \text{and} ^{31}\text{P}\). In practice, hydrogen nuclei, \(i.e. \(^1\text{H}\), are used for signal generation in MRI due to their abundance in the living body and their magnetic properties (see Table A.1, data courtesy Panchuelo \([106]\)). A hydrogen nuclear is composed of only one proton; in this chapter, proton indicates the proton in the hydrogen nuclear.

Table A.1: Magnetic properties of nuclei.

<table>
<thead>
<tr>
<th>Nucleus</th>
<th>Spin</th>
<th>(\gamma) (MHz/T)</th>
<th>Relative sensitivity (to (^1\text{H}))</th>
<th>Natural abundance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(^1\text{H})</td>
<td>1/2</td>
<td>42.58</td>
<td>1</td>
<td>99.8</td>
</tr>
<tr>
<td>(^{13}\text{C})</td>
<td>1/2</td>
<td>10.71</td>
<td>1.59 \times 10^{-2}</td>
<td>1.1</td>
</tr>
<tr>
<td>(^{14}\text{N})</td>
<td>1</td>
<td>3.08</td>
<td>1.01 \times 10^{-2}</td>
<td>99.6</td>
</tr>
<tr>
<td>(^{23}\text{Na})</td>
<td>3/2</td>
<td>11.26</td>
<td>9.25 \times 10^{-2}</td>
<td>100</td>
</tr>
<tr>
<td>(^{31}\text{P})</td>
<td>1/2</td>
<td>17.24</td>
<td>6.63 \times 10^{-2}</td>
<td>100</td>
</tr>
</tbody>
</table>

Protons in nuclei always spin on their axes. Their spins produce a magnetic field called magnetic moment. In the normal state, spin axes of protons are oriented randomly. The spins cancel each other out; thus, there is no overall magnetic moment. Fig. A.1 illustrates a proton spinning on its axis and protons aligned randomly in the absence of a magnetic field.

![Figure A.1](image1.png)

**Figure A.1:** An illustration of a proton spinning on its axis and protons aligned randomly in the absence of a magnetic field.

When placed in an external magnetic field, \(e.g.\) in the magnetic field of an MRI scanner, protons spin and precess on their axes like a spinning top (see Fig. A.2(a)).
The precession axis is parallel to the direction of the external magnetic field. The precession rate (cycles per second) is determined by the Larmor frequency \( f = \gamma B_0 / 2\pi \), where \( \gamma \) is the gyromagnetic ratio. For hydrogen nuclei, the gyromagnetic ratio is \( \gamma = 42.56\text{MHzT}^{-1} \). The precession rate is called as resonant frequency or Larmor frequency.

There are two states of spins when they are precessing: parallel and anti-parallel (see Fig. A.2(b)). Spins in the parallel state have a lower energy level and are more stable than protons in the anti-parallel state. The proportion of spins in the two states depends on the temperature and strength of the external magnetic field. In a low temperature and a high magnetic field, there are more spins in the parallel state than in the anti-parallel state. In Earth’s magnetic field at room temperature, roughly equal numbers of protons are in the two energy states, with only slightly more in the parallel state.

Adding up spins together results in a net magnetization. Since spins are opposite along the magnetic field, there is no net magnetization in the direction perpendicular to the magnetic field. The net magnetization is parallel to the magnetic field; thus, it is known as the longitudinal magnetization. Fig. A.3 illustrates the composition of the longitudinal magnetization from the spins placed in an external magnetic field. The magnitude of the longitudinal magnetization depends
A.1. MR signal generation

on the difference between the number of spins in the parallel state and the number of spins in the anti-parallel state.

![Diagram of net magnetization](image)

**Figure A.3:** An example of the net magnetization.

In a homogeneous magnetic field, where the magnetic strength is the same at every point in the field, the net magnetization in a volume of space depends on the number of protons. Because the distribution of hydrogen atoms are varied in different tissues, the magnitude of the net magnetization provides a representation of the body structure. The MR technique is based on measuring the net magnetization to image the tissues inside a living body. Measuring the net magnetization can not be done in the equilibrium condition. The measuring can be done by perturbing the equilibrium condition. This operation is known as **excitation** and **relaxation**.

**Excitation** is started by continuously applying radio-frequency (RF) pulses with a frequency corresponding to the Larmor frequency to the vicinity of protons (Fig. A.4(a)). Spins in the low-energy state will absorb the energy and turn into the high-energy state. In MRI, the amount of electromagnetic energy is enough to generate equal numbers of spins in each state. The longitudinal magnetization is gradually decreased to zero until the numbers of spins in each state are equal. Spins after this step is known as **out of phase** (Fig. A.4(b)).

The sinusoidal RF pulses then push spins to synchronize and precess together. This is the resonance portion of the NMR. As a result, the net magnetization turns towards the transverse plane, which is perpendicular to the magnetic field. The net magnetization at this stage is known as the **transverse magnetization**. Spins after this step is known as **in phase** (Fig. A.4(c)).

**Relaxation** is started by turning the RF pulses off. The spin system, which
A.1. MR signal generation

Figure A.4: States of spins during the excitation.

is in an excited state, now turns into the equilibrium condition. Spins being all positively charged will repel each other and move apart. As they spread apart, the transverse magnetization is decreased over time, and finally disappears. The spin system returns to the state like Fig. A.4(b). The other relaxation occurs as the high-energy spin shift to the low-energy state. As these spin switch to the baseline, the longitudinal magnetization is regrown as in Fig. A.3. During that process, the transverse magnetization quickly loses coherence, and the longitudinal magnetization slowly recovers.

The changes in the net magnetization over time is measurable via a receiver coil. The received signal is called MR signal; it is used to construct the image of the scanned subject. The decay in transverse magnetization is called transverse ($T_1$) relaxation, while the recovery of longitudinal magnetization is called longitudinal ($T_2$) relaxation. By specifying a pulse sequence that targets one of these relaxation parameters, images can be collected that are sensitive to specific properties of the underlying tissue.
A.2 MR image formation

This section presents an overview of reconstructing the image from received MR signals. The fundamental concept of image formation in MRI is the use of magnetic gradients, which is used to provide the spatial contributions to the MR signals [107]. The idea is that a gradient field, which varies linearly along a particular axis, is introduced during the excitation step. This application of the gradient field results in a linear variation of the resonant frequency along that axis. In this way, the received MR signals are the summation of components with different frequencies bring the net magnetizations over time and their spatial information. Thus, it is possible to generate the images that provide information about the characteristic of those atomic nuclei.

In detail, we look at the decomposition of the net magnetization $M$ along the x-axis, y-axis, and z-axis, denoted as $M_x$, $M_y$, and $M_z$, respectively. The changes of the three magnetization components over time are described using Block equations:

$$
\frac{dM_x}{dt} = \gamma B_t M_y - \frac{M_x}{T_2}, \quad \frac{dM_y}{dt} = \gamma B_t M_x - \frac{M_y}{T_2}, \quad \frac{dM_z}{dt} = -\frac{M_z - M_0}{T_1}, \quad (A.1)
$$

where $B_t$ is the total magnetic field; $M_0$ is the equilibrium magnetization; $T_1$ denotes the time constant associated with the recovery of the longitudinal magnetization; and $T_2$ denotes the time constant associated with the decay of the transverse magnetization [16]. In the presentation of the field gradients, $B_t = B_0 + G_x x + G_y y + G_z z$, where $B_0$ is the strength of the external magnetic field, $G_x$, $G_y$, and $G_z$ denote the gradient field along the x-axis, y-axis, and z-axis, respectively. Note that the z-axis known as the longitudinal axis is parallel to the main magnetic field. The x-y plane known as the transverse plane is perpendicular to the main magnetic field.

The transverse magnetization, denoted by $M_{xy}$, can be represented by the $M_x$ and $M_y$ components. The transverse magnetization $M_{xy}(x, y, z, t)$ for a given
spatial location and time instance is

\[ M_{xy}(x, y, z, t) = M_{xy0}(x, y, z) e^{-t/T_2} e^{-i\omega t}, \quad (A.2) \]

where the term \( M_{xy0}(x, y, z) \) denotes the original magnetization (time \( t = 0 \)) at the spatial location \((x, y, z)\). The term \( \omega \) is the angular frequency of the transverse magnetization in the total magnetic field \( B_t \) which is followed the Larmor frequency equation \( \omega = 2\pi f = \gamma B_t \). The detail representation of Eq. (A.2) is

\[ M_{xy}(x, y, z, t) = M_{xy0}(x, y, z) e^{-t/T_2} e^{-i\gamma B_0 t} e^{-i\int_0^t (G_x(\tau)x + G_y(\tau)y + G_z(\tau)z) d\tau}. \quad (A.3) \]

The MR signal measured by the receiver coil reflects the sum of the transverse magnetization of all voxels within the excited sample. The signal can be expressed as

\[ S(t) = \iiint M_{xy}(x, y, z, t) \, dx \, dy \, dz. \quad (A.4) \]

Combining Eqs. (A.3) and (A.4), we get the signal measured at any time instance \( t \) as

\[ S(t) = \iiint M_{xy0}(x, y, z) e^{-i\gamma \int_0^t (G_x(\tau)x + G_y(\tau)y + G_z(\tau)z) d\tau} \, dx \, dy \, dz. \quad (A.5) \]

As shown in [108], Eq. (A.5) can be simplified as

\[ S(t) = \iiint M_{xy0}(x, y, z) e^{-i\int_0^t (G_x(\tau)x + G_y(\tau)y + G_z(\tau)z) d\tau} \, dx \, dy \, dz. \quad (A.6) \]

This equation illustrates the profound importance of the gradient fields for encoding spatial information within an MR image. It has the form of the Fourier transform. Thus, the image can be reconstructed by an inverse Fourier transform. The received signals are sampled and stored in special storage known as \textit{k-space}. Fig. 2.1 demonstrates the flow of data in the MR image formation.

In practice, we can collect the MR signal of an entire volume (known as a 3D imaging technique) by systematically turning on gradient fields along \( x, y, \) and \( z \).
A.2. MR image formation

directions. The 3D imaging sequence presents additional technical challenges and
is less tolerant of hardware imperfection. A more common approach to collect
the emitted signal is the 2D imaging technique, which reduces the signal equation
into two dimensions.

The 2D MR imaging technique forms 3D images from sets of slices. A series of
RF pulses and field gradients is applied to generate a slice. First, a field gradient is
introduced along a direction. This causes spins precessing at different frequencies
along the direction. Next, an RF pulse is first applied to the vicinity of protons.
The frequency band of the RF pulse is the same as the frequency band of the
spins in the selected slice; thus, all spins within the selected slice are excited. This
gradient application is called slice selection.

Once a slice is selected, all excited spins contribute equally to the generation
of the MR signal, which contains no spatial information at this point in the pulse
sequence. Thus, the next step is to apply additional gradients that cause spins
at different spatial locations to precess at different rates in a controlled manner,
so that their individual contributions can be measured. For this reason, phase-
encoding gradients and frequency-encoding gradients are in turn applied. The
phase-encoding gradient is used to forced spins in the same row precessing with
the same phase. The frequency-encoding gradient is used to forced spins in the
same column precessing the same frequency.

The 3D MR imaging technique collects k-space data in three dimensions;
thus, the 3D images are obtained via the inverse Fourier transform of the k-space
data directly. In 3D imaging technique, the traditional slice excitation step is
replaced by a volume excitation step that uses a very small gradient field to select
a thick slice. This step is similar to applying a phase-encoding gradient along that
dimension. Therefore, a typical 3D pulse sequence contains two phase-encoding
gradients and one frequency-encoding gradient.
Appendix B

Multi-modal similarity measures

Modality independent neighborhood descriptor (MIND) is a multi-dimensional descriptor, which was proposed for computing the dissimilarity measure in multi-modal deformable image registration [109]. This descriptor is independent of the modality, contrast, and noise level of images since it captures the self-similarity of the image patches around a voxel.

The multi-dimensional descriptor $s_{\text{MIND}}$ of a voxel $x$ within the search space $R$ (centered at $x$) is a vector with the length as the number of elements in $R$. The MIND value of a voxel of $x$ in an image $I$ at a single entry $r_i \in R$ is defined as

$$s_{\text{MIND}}(I, x, r_i) = \frac{1}{n} \exp\left(-\frac{d_l(I, x, r_i)}{v(I, x)}\right),$$

where $n$ is a constant to normalize the maximum value as 1. Here, $d_l(I, x_1, x_2) = \sum_{p \in \mathcal{P}} G_o(x_1 + p)(I(x_1 + p) - I(x_2 + p))^2$ is the patch-based dissimilarity, $\mathcal{P}$ denotes neighborhood indexes of a patch size of $(2l+1)$. The term $G_o$ denotes the image obtained by applying a Gaussian filter (having the same kernel size as the patch) on the difference image between image $I$ and its shifted version from $x_1$ to $x_2$. The term $v(I, x)$ is the mean patch-based dissimilarity of voxel $x$ in image $I$ with its six neighbors.
The MIND-based dissimilarity of images $A$ and $B$ is defined as the sum of MIND difference at every voxel in the image domain $\Omega$:

$$D_{\text{MIND}}(A, B) = \frac{1}{|\Omega|} \sum_{x \in \Omega} D_{\text{MIND}}(A, B, x),$$  \hspace{1cm} (B.2)

where

$$D_{\text{MIND}}(A, B, x) = \frac{1}{|R|} \sum_{r_i \in R} |s_{\text{MIND}}(A, x, r_i) - s_{\text{MIND}}(B, x, r_i)|.$$  \hspace{1cm} (B.3)

A large value of the $D_{\text{MIND}}$ indicates the more structural dissimilarity between the two images $A$ and $B$.

Fig. B.1 shows examples of MIND difference maps for different scenarios. The first is the MIND map between $T_{1w}$ images of two different subjects, see Fig. B.1 (d). The second is the MIND map between the $T_{1w}$ image and the EPI image of the same subject, see Fig. B.1 (e). It can be seen that the MIND-based dissimilarity measure between different subjects ($T_{1w}$ to $T_{1w}$) is larger than the MIND-based dissimilarity measure between different image modalities of the same subject ($T_{1w}$ to EPI).
Figure B.1: An example of MIND difference maps between T1w and other MRI types: (a) T1w image of Subject 1; (b) T1w image of Subject 2; (c) EPI image of Subject 1; (d) MIND map of (a) and (b) with MIND score $D = 0.42$; (e) MIND map of (a) and (c) with MIND score $D = 0.24$. A blue color denotes a small difference, a red color denotes a large difference. See the electronic color image.
Bibliography


