A DEEP LEARNING MODEL FOR SUSCEPTIBILITY ARTIFACT CORRECTION IN ECHO PLANAR IMAGING

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF SCIENCE

IN

ELECTRICAL AND ELECTRONICS ENGINEERING

By

Abdallah Ghazi Faisal Zaid Alkilani

September 2022

A DEEP LEARNING MODEL FOR SUSCEPTIBILITY ARTIFACT CORRECTION IN ECHO PLANAR IMAGING
By Abdallah Ghazi Faisal Zaid Alkilani
September 2022

We certify that we have read this thesis and that in our opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Emine Ülkü Sarıtaş Çukur (Advisor)

Tolga Çukur (Co-Advisor)

Ergin Atalar

Aydın Akan

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

A DEEP LEARNING MODEL FOR SUSCEPTIBILITY ARTIFACT CORRECTION IN ECHO PLANAR IMAGING

Abdallah Ghazi Faisal Zaid Alkilani
M.S. in Electrical and Electronics Engineering
Advisor: Emine Ülkü Sarıtaş Çukur
Co-Advisor: Tolga Çukur
September 2022

Diffusion Weighted Imaging (DWI) is a Magnetic Resonance Imaging (MRI) technique that probes the Brownian motion of water molecules within biological tissue, *in vivo* and non-invasively. The most commonly employed sequence for DWI is Echo Planar Imaging (EPI), where the entirety of k-space is collected in a zigzag manner in one shot after a single diffusion preparation step. EPI is preferred due to its robustness to motion, and it meets the high signal-to-noise ratio efficiency and short acquisition duration demands of DWI. However, EPI suffers from severe susceptibility-induced artifacts that arise due to strong readout gradients and low bandwidth along the phase-encode (PE) direction. These artifacts are caused by magnetic susceptibility differences and manifest as geometric and intensity distortions. Postprocessing methods are extensively used to correct for these artifacts, particularly reversed PE techniques that utilize images acquired in reversed PE directions to deduce the susceptibility-induced displacement field. While many non-learning methods exist for the reversed PE approach, they are relatively time consuming and require instance-specific optimization. Only a few recent works have explored the benefits of employing deep learning to speed up the reversed PE approach. These methods rely on unwarping correction with a predicted displacement field that maximizes image similarity.

This thesis proposes a deep unsupervised Forward-Distortion Network (FD-Net) for correcting susceptibility artifacts. FD-Net speeds up the correction while explicitly constraining measurement fidelity for enhanced correction performance. This technique employs an encoder-decoder architecture to predict the field as well as the corrected image from the input reversed PE images. Using the field to forward-distort the predicted image in both PE directions should explain the

input reversed PE images, thereby enforcing consistency to input data. This forward-distortion approach relies on matrix operations and is computationally efficient. Two different multiresolution strategies are considered: a multiscale strategy where earlier stages of the decoder produce lower resolution field and image predictions, and a multiblur strategy where the full resolution predictions are progressively blurred. Both strategies aim to boost performance by enforcing consistency across different scales and blurs, effectively speeding up convergence and circumventing local minima. In this thesis, variations of the multiresolution strategies are considered and the highest performing strategy in terms of quantitative image quality metrics is chosen.

The performance of FD-Net is evaluated in comparison to two recent deep learning methods from the literature and a supervised baseline method based on FD-Net. A classical unwarping-based method is used as the gold standard reference. Extensive slice-wise, subject-wise, visual, and quantitative image quality assessments are performed. The results demonstrate that FD-Net surpasses the competing deep learning methods, and outperforms the supervised baseline in terms of predicted image quality, while maintaining robust field predictions. Hence, the forward-distortion model presents a better-conditioned problem for distortion correction when compared to unwarping-based approaches. This thesis concludes that FD-Net provides a novel paradigm for the susceptibility artifact correction problem that better constrains fidelity to the measurement data.

Keywords: Susceptibility artifacts, echo planar imaging, reversed phase-encoding, deep learning, unsupervised learning.

ÖZET

EKO PLANAR GÖRÜNTÜLEMEDE DUYARLILIK ARTEFAKTI DÜZELTME İÇİN DERİN ÖĞRENME MODELİ

Abdallah Ghazi Faisal Zaid Alkilani
Elektrik ve Elektronik Mühendisliği, Yüksek Lisans
Tez Danışmanı: Emine Ülkü Sarıtaş Çukur
İkinci Tez Danışmanı: Tolga Çukur
Eylül 2022

Difüzyon ağırlıklı görüntüleme (DWI), biyolojik doku içindeki su moleküllerinin Brown hareketini *in vivo* ve noninvaziv şekilde incelemeyi sağlayan bir manyetik resonans görüntüleme (MRI) tekniğidir. DWI için en yaygın olarak kullanılan sekans, tek bir difüzyon hazırlama adımından sonra tek seferde k-uzayının tamamının zikzak şeklinde toplandığı eko planar görüntülemedir (EPI). EPI, harekete karşı gürbüzlüğü nedeniyle tercih edilmektedir ve DWI'nın yüksek sinyal-gürültü oranı verimliliği ve kısa alım süresi gereksinimlerini karşılamaktadır. Bunlara rağmen, EPI, güçlü okuma gradyanları ve faz-kodlama (PE) yönündeki düşük bant genişliği nedeniyle ortaya çıkan duyarlılık artefaktlerinden muzdariptir. Bu artefaktlar, manyetik duyarlılık farklılıklarından kaynaklanmaktadır, ve geometrik bozulma ve yoğunluk bozulması olarak kendini göstermektedir. Bu artefaktları düzeltmek için son işlem yöntemleri yaygın olarak kullanılmaktadır. Bu yöntemler arasında özellikle duyarlılığa bağlı yer değiştirme alanını belirlemek için ters PE yönlerinde elde edilen görüntüleri kullanan ters PE teknikleri yaygındır. Ters PE yöntemleri için birçok derin öğrenme dışı yöntem mevcut olsa da, bunlar nispeten zaman alıcıdır ve örneğe özgü optimizasyon gerektirmektedir. Yalnızca birkaç yeni çalışma, ters PE tekniklerini hızlandırmak için derin öğrenmeyi kullanmanın faydalarını araştırmıştır. Bu yöntemler görüntü benzerliğini en üst düzeye çıkaran tahmini bir yer değiştirme alanıyla tersine çarpıtarak düzeltmeye dayanmaktadır.

Bu tez, duyarlılık artefaktlarını düzeltmek için derin ve denetimsiz bir ileri yönde bozulma ağı (FD-Net) önermektedir. FD-Net, gelişmiş düzeltme performansı için ölçüm doğruluğunu açıkça kısıtlarken düzeltmeyi hızlandırmaktadır.

Bu teknik, girdi ters PE görüntülerinden alanı ve düzeltilmiş görüntüyü tahmin etmek için bir kodlayıcı-kod çözücü mimarisi kullanmaktadır. Tahmin edilen görüntüye her iki PE yönünde ileri yönde bozulma işlemi gerçekleştirmek için alanın kullanılması, girdi ters PE görüntülerini açıklamalı ve böylece girdi verilerine dayanarak tutarlılığı sağlamalıdır. Bu ileri yönde bozulma yaklaşımı, matris işlemlerine dayanmaktadır ve hesaplama açısından verimlidir. İki farklı çoklu çözünürlük stratejisi göz önünde bulundurulmuştur: Kod çözücünün önceki aşamalarının daha düşük çözünürlüklü alan ve görüntü tahminleri ürettiği bir çoklu ölçek stratejisi, ve tam çözünürlüklü tahminlerin aşamalı olarak bulanıklaştığı bir çoklu bulanıklık stratejisi. Her iki strateji de farklı ölçekler ve bulanıklıklar arasında tutarlılığı zorunlu kılarak performansı artırmayı amaçlamakta, böylece etkin bir şekilde yakınsamayı hızlandırmakta ve yerel minimumlardan kaçınmaktadır. Bu tezde çoklu çözünürlüklü stratejilerinin varyasyonları göz önünde bulundurulmaktadır, ve nicel görüntü kalitesi metrikleri açısından en yüksek performans gösteren strateji seçilmektedir.

FD-Net'in performansı, literatürdeki iki yeni derin öğrenme yöntemiyle ve FD-Net'e dayanan bir denetimli temel yöntemle karşılaştırılarak değerlendirilmektedir. Altın standart referans olarak klasik bir tersine çarpıtma tabanlı yöntem kullanılmaktadır. Görüntü kesiti bazında, katılımcı bazında, görsel ve nicel görüntü kalitesi değerlendirmeleri kapsamlı olarak yapılmaktadır. Sonuçlar, FD-Net'in karşılaştırılan derin öğrenme yöntemlerinden daha iyi performans gösterdiğini, ve gürbüz alan tahminlerini sürdürürken, tahmin edilen görüntü kalitesi açısından denetimli temel yöntemden üstün olduğunu göstermektedir. Bu nedenle, ileri yönde bozulma modeli, tersine çarpıtma tabanlı yaklaşımlarla karşılaştırıldığında, bozulma düzeltmesi için daha iyi koşullandırılmış bir problemi sunmaktadır. Bu tez, FD-Net'in ölçüm verilerinin aslına uygunluğunu daha iyi sınırlayan duyarlılık artefaktı düzeltme problemi için yeni bir yaklaşım sağladığı sonucuna varmaktadır.

Anahtar sözcükler: Duyarlılık artefaktları, eko planar görüntüleme, ters faz kodlaması, derin öğrenme, denetimsiz öğrenme.

Acknowledgement

First of all, I would like to express my gratitude to my advisor, Assoc. Prof. Emine Ülkü Sarıtaş, for her support throughout the duration of my master's studies. Her extensive experience, excellent teaching skills, along with her patience and supportive guidance, motivated me to push forward in my academic life.

I would like to thank Assoc. Prof. Tolga Çukur, for his invaluable insights and suggestions in deep learning and for co-advising this thesis.

I would also like to thank Prof. Ergin Atalar and Prof. Aydın Akan for joining my thesis committee. Their invaluable feedback helped make this thesis more complete.

I am thankful to the Scientific and Technological Council of Turkey for supporting the work in this thesis through TÜBİTAK Grant 117E116.

I am also thankful to the Human Connectome Project and the McDonnell Center for Systems Neuroscience at Washington University, for providing the data.

I would like to thank my dear friends Muntaha, Tina, Alper, and Amir for their endless support in my personal life. I would also like to thank Yusuph and Muhammad for their camaraderie during my master's studies.

Last but not least, I would like to thank my family. My parents have always encouraged and appreciated the value of a good education, providing loads of love and support during my time as a student. My brother Faisal was always the cool but caring older sibling. My little sister, Sarah, has been my friend for what feels like an eternity now, is always there to listen to my rants and to cheer me up with cat photos. I am truly blessed to have such a strong bond with her.

I dedicate this thesis in loving memory of Uncle Muammar.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Echo planar imaging (EPI) [1] is the most commonly employed Magnetic Resonance Imaging (MRI) sequence for Diffusion Weighted Imaging (DWI) and functional Magnetic Resonance Imaging (fMRI), due to its rapid k-space acquisition capability [2, 3]. However, EPI is prone to susceptibility artifacts arising from B_0 field inhomogeneities due to the magnetization of the object being imaged, which is problematic especially at tissue-to-tissue and tissue-to-air interfaces [4]. These artifacts materialize as intensity distortions from signal pileups/dropouts, and geometrical distortions due to compression/stretching of affected regions [5]. The severity of these artifacts can reduce the clinical utility of the acquired images and necessitate correction for further quantitative analyses and interpretations [6], especially in the case of high-field MRI [7, 8].

Leading methods for susceptibility artifact correction utilize images acquired in reversed phase-encoding (PE) directions to estimate the susceptibility-induced displacement field directly from the resulting blip-up/blip-down EPI images [5, 9, 10, 11]. These methods have been shown to outperform multiple-echo measured-field techniques, as well as registration-based techniques, which were shown to exhibit poor performance [12]. Correction is commonly done by unwarping the reversed PE images based on an estimated displacement field. This field can either be represented as pixel-wise or voxel-wise parameters that allows for increased freedom in capturing the spatial changes in the field [13, 14, 10], or as weighting coefficients to linear basis functions [9]. Popular implementations of this reversed PE method include TOPUP from the FMRIB Software Library (FSL) [15, 9] and hyperelastic susceptibility correction of DTI data (HySCO) of the Statistical Parametric Mapping (SPM) toolbox [16, 17]. Other correction approaches include point spread function (PSF) based approaches [18], which require additional scans and/or further assumptions on the acquired data (such as the k-space trajectory).

Even when the off-resonance frequency at each pixel/voxel location is known, it is nontrivial to determine the underlying displacement field, due in part to the complex interactions at the tissue-to-tissue and tissue-to-air interfaces [19]. This difficulty emphasizes the need to estimate the field from images with identical contrast. More concretely, unwarping-based methods attempt to find an "average space" based on the observations to maximize the similarity of the corrected blip-up/blip-down EPI images, which can lead to changes in noise properties and biased parameter estimates. This problem can be mitigated by utilizing a data modelling framework, where a model is built from the training data, together with its parameters and a means to make predictions about what the underlying correct data should look like by comparison with the distorted observation. This process can be thought of as transforming the model to the space of the observation [20].

Deep learning approaches in recent literature gravitate towards the unwarping-based method with a convolutional encoder-decoder U-Net [21] architecture that predicts the displacement field from the input reversed PE images. S-Net operates on 3D volumes and performs unwarping with a bilinear interpolation operation, using a multimodal loss to assess image similarity [22]. However, field smoothness requirements in S-Net can cause insufficient intensity correction when large distortions and pileups are present, leading to underperformance in terms of image correction. Deepflow-Net operates on 2D slices and uses cubic interpolation for unwarping, using mean squared error (MSE) as an image similarity loss together with a density compensation scheme and a multiresolution strategy to boost the correction performance [23]. Even though Deepflow-Net addresses the issue of insufficient correction by using a density compensation scheme, it relies on the inversion of a density pileup map. Therefore, it introduces issues when modulating large distortions, leading to a less perceptively smooth field and manifesting as intensity perturbations in the corrected image. Despite potential shortcomings, these deep learning based approaches provide a considerable boost to the speed of correction.

This thesis proposes a deep Forward-Distortion Network (FD-Net) for correcting susceptibility artifacts by predicting the field and the underlying corrected image. Forward-distorting the corrected image with the field explains input distorted images by facilitating comparisons directly in the observation space. FD-Net builds on the concept of "amortized optimization" or "amortization", were acquisition-specific optimization is replaced by a global optimization of shared parameters [24]. FD-Net is validated with respect to TOPUP and the performance boost from combining the forward-distortion approach with a multiresolution scheme is investigated. The superiority of FD-Net over prior unwarping-based deep learning methods as well as a supervised baseline implementation is demonstrated, thereby establishing FD-Net as a quick and effective model for susceptibility artifact correction in EPI.

Chapter 2

Background

EPI is a popular imaging sequence with a variety of applications in domains such as diffusion MRI, fMRI, cardiac imaging, and perfusion imaging [25]. With its capability to acquire the entire k-space at once, EPI is one of the fastest data acquisition techniques in MRI.

This chapter presents the motivation behind using EPI for one of these applications, namely diffusion MRI. First, some background on diffusion MRI is provided. Then, EPI is introduced and its usefulness in diffusion MRI is emphasized, leading to a discussion of the artifacts present in EPI with a particular focus on the susceptibility artifacts. Finally, the existing distortion correction methods for susceptibility artifacts are discussed.

2.1 Diffusion MRI

Diffusion MRI (also called Diffusion Weighted MRI or DWI) is a unique technique for probing diffusion in biological tissue, *in vivo* and non-invasively, and without disturbing the diffusion process [26]. Whether in a normal or in a diseased state, DWI reveals details about tissue microstructure by resolving the diffusion pattern of water molecules.

DWI is used in a variety of clinical applications such as diagnosing acute ischemic stroke [27, 28, 29], lymphadenopathy [30], liver tumors [31], renal masses [32, 33], prostate cancer [34], and colorectal cancer [35]. DWI is also popular in the field of neuroscience due to its utility for investigating the microstructure of white matter and for fiber tracking/tractography [36, 37, 38].

2.1.1 Diffusion in Biological Tissue

In the absence of flow or external agitation, water molecules experience random microscopic displacements due to internal thermal energy and undergo Brownian motion. Due to the random nature of these displacements, the ensemble average of the displacement for a group of water molecules is zero, with a Gaussian probability density function (PDF) [39]. The variance of the displacement r can be characterized in n dimensions as a function of elapsed time t as

$$\langle r^2 \rangle = 2nDt, \tag{2.1}$$

where D is the diffusion coefficient described via Fick's first and second laws of diffusion [40] expressed in units of area over time (e.g., mm^2/s), and n is the number of spatial dimensions. Equation (2.1) is a generalization to n dimensions of the Einstein-Smoluchowski equation, which describes unhindered diffusion in a single dimension.

Equation (2.1) characterizes diffusion in free water, which is isotropic. The Stokes-Einstein theorem postulates that isotropic diffusion relates to the environment. Hence, information on the physical properties of the environment in which water molecules experience isotropic diffusion is provided via the diffusion coefficient, D. However, diffusion is often anisotropic in biological tissue, particularly the nervous tissue. The presence of cell membranes and microscopic structures leads water molecules to experience hindrance and restriction in the extracellular and intracellular spaces, respectively [41]. The complex microstructure of tissue can be delineated by modelling diffusion through one of the many models available in the literature, from the simple tensor model [26] to more complex models such as Neurite Orientation Dispersion and Density Imaging (NODDI) [42].

These models enable the extraction of useful biomarkers for tissue structure and integrity.

2.1.2 Diffusion Contrast in MRI

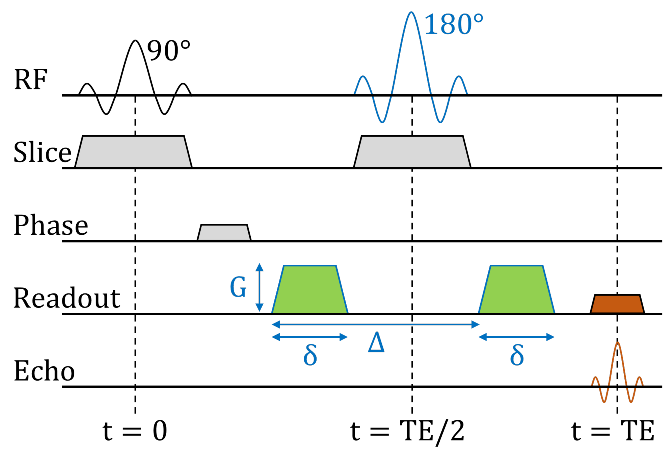

Figure 2.1: Illustration of a simple pulsed gradient SE sequence. After slice selection using the 90° RF pulse, the diffusion sensitization process is started (colored in blue): two bulk diffusion-sensitizing gradients of amplitude G (shown in green) are placed around the 180° refocusing RF pulse. The durations of these gradients are denoted by δ and their separation is denoted by Δ. The signal readout is centered around TE (shown in brown). In this example, the diffusion weighting is applied along the readout direction.

The idea of using a Spin Echo (SE) sequence to measure diffusion was first proposed c. 1950 [43, 44], with considerable improvements introduced by the pulsed field gradient method [45]. Figure 2.1 illustrates how this method employs a pair of magnetic field gradients placed before and after the refocusing radio frequency (RF) pulse to spatially encode the random movement of the spins. Accordingly, the spins accumulate a phase ϕ by the echo time (TE), which can be characterized for a single spin as

$$\phi(TE) = \gamma \left[\int_0^{\frac{TE}{2}} \mathbf{G}(t) \cdot \mathbf{x}(t)\, dt - \int_{\frac{TE}{2}}^{TE} \mathbf{G}(t) \cdot \mathbf{x}(t)\, dt \right], \tag{2.2}$$

where γ is the gyromagnetic ratio, $\mathbf{G}(t)$ is the diffusion-sensitizing gradient field vector, and $\mathbf{x}(t)$ is the spatial coordinate vector. Identical phase is incurred on coherently moving spins, while zero phase is incurred on stationary spins (i.e., the integrals from Equation (2.2) cancel out). Diffusion, on the other hand, incurs a different phase amount to each spin due to the random motion experienced under each gradient lobe (i.e., $\mathbf{x}(t)$ from Equation (2.2) evolves in a random fashion).

The PDF of the phase $\phi(TE)$ can be shown to be a Gaussian with zero mean and variance σ_ϕ^2, since the displacement denoted by $\mathbf{x}(t)$ from Equation (2.2) is also a Gaussian [44] (cf. Subsection 2.1.1). The diffusion-weighted signal intensity can be obtained relative to the non-diffusion-weighted signal intensity by integrating $\phi(TE)$ for an ensemble of spins, i.e.,

$$S = S_0 \int_{-\infty}^{\infty} f_\phi e^{i\phi}\, d\phi = S_0\; e^{-\frac{\sigma_\phi^2}{2}} = S_0\; e^{-bD}, \tag{2.3}$$

where f_ϕ is the PDF of $\phi(TE)$, S_0 is the signal in the absence of diffusion-sensitizing gradients, and b refers to the so-called "b-value" (expressed in units of time over area, e.g., s/mm^2). The Gaussian nature of the phase dispersion leads to destructive interference that is exhibited as an exponential signal attenuation, but with zero net phase shift [46].

Together with the diffusion coefficient D, the b-value controls the extent to which diffusion weighting is applied. The b-value depends on the applied diffusion gradient strength as well as its duration, and can be expressed at $t = TE$ in its general form as [46]

$$b = \gamma^2 \int_0^{TE} \left| \int_0^t \hat{\mathbf{G}}(\tau)\, d\tau \right|^2 dt, \tag{2.4}$$

where

$$\hat{\mathbf{G}}(t) = \begin{cases} \mathbf{G}(t), & t < \frac{TE}{2}, \\ -\mathbf{G}(t), & t \geq \frac{TE}{2}. \end{cases} \tag{2.5}$$

When $b = 0$, no diffusion sensitization is involved and the corresponding acquisition is used to determine S_0 (cf. Equation (2.3)). Images collected with one or more non-zero b-values in one or more diffusion-sensitizing directions, together with one or more $b = 0$ (or $b \approx 0$) images, can be used together with a suitable model to extract high level details about the underlying microstructure (cf. Subsection 2.1.1). Examples of DWI in the brain are shown in Figure 2.2.

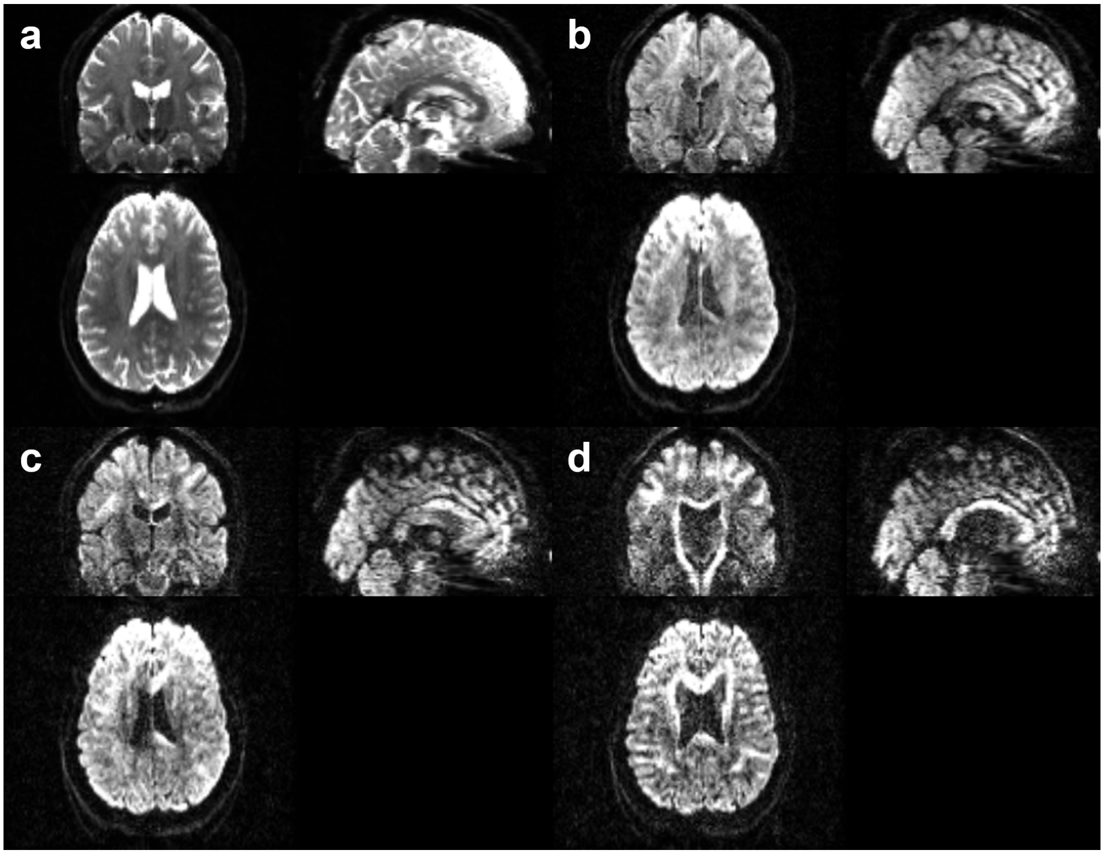

Figure 2.2: Examples of DWI in the brain with different b-values. For each case, from top left, clockwise: coronal, sagittal, and axial views. The same cross-sections are displayed for (a) $b = 5\ s/mm^2$ (effectively $b = 0$), (b) $b = 1000\ s/mm^2$, (c) $b = 1800\ s/mm^2$, and (d) $b = 2500\ s/mm^2$. Data was acquired in the National Magnetic Resonance Research Center (UMRAM) on a 3T Siemens MAGNETOM Trio scanner with a monopolar diffusion scheme using single-shot EPI readout with posterior-to-anterior PE polarity. Distortion correction was implemented in postprocessing. Other imaging parameters were $260 \times 225\ mm^2$ field-of-view (FOV), $2\ mm$ isotropic resolution, averages = 1, multi-band acceleration factor 4; TR/TE = $3510/114.4\ ms$, flip angle = $78°$; 132×114 acquisition matrix, bandwidth = $1403\ Hz/Px$, EPI factor = 114, echo spacing = $0.97\ \mu s$, and 6/8 phase partial Fourier acquisition.

2.2 EPI Sequence for Diffusion MRI

As discussed in Subsection 2.1.2, diffusion contrast is based on attenuation of the MRI signal. Diffusion gradients sensitize the sequence to micrometer level motion. In the presence of global rigid motion during these gradients, the k-space data suffers from additional phase issues and a k-space shift problem, reducing the utility of otherwise useful conventional sequences such as line-by-line imaging [47]. Examples of this global rigid motion include patient movements such as head motion and breathing [48]. Additionally, multiple acquisitions with differing b-values and/or diffusion directions are typically required for microstructural models, and acquiring multiple slices benefit fiber tracking applications [49]. These requirements demand high signal-to-noise ratio (SNR) efficiency, while preserving a short acquisition duration for clinical utility [47].

By far, the most popular sequence that addresses these requirements for DWI is EPI [1]. EPI is a multi-planar image formation sequence that allows the traversal of k-space in a zigzag manner, compared to line-by-line Fourier acquisition in a typical sequence.

In single-shot EPI (ss-EPI), the entirety of k-space is collected following a single diffusion preparation step with a single excitation, making ss-EPI robust to motion during diffusion-sensitizing gradients [50]. The acquisition matrix in ss-EPI typically does not exceed 128×128, making the resolution limited. Hence, in order to image at a higher resolution, multi-shot EPI (ms-EPI) can be performed [51], such as interleaved EPI [52]. ss-EPI still remains the *de facto* choice for DWI, since the motion-induced phase differences and k-space shifts among the different shots in ms-EPI can lead to ghosting artifacts and signal voids [47], the correction of which requires acquiring additional navigator echoes [53]. Figure 2.3 illustrates the differences in k-space acquisition between line-by-line acquisition, ss-EPI, and ms-EPI.

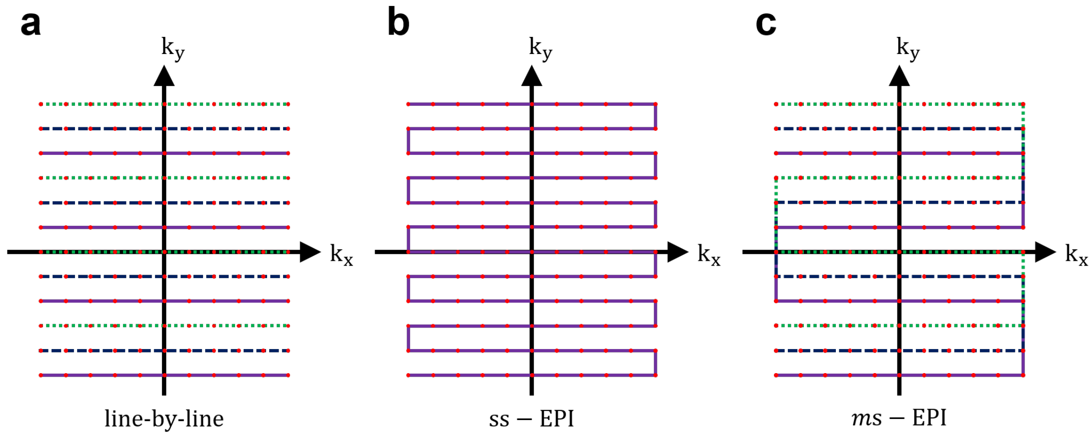

Figure 2.3: Illustration of k-space for different data acquisition strategies. (a) In a typical sequence, k-space is acquired line-by-line with a separate excitation for each line. (b) In ss-EPI, the entirety of k-space is acquired in one shot after a single excitation. Traversal of k-space is performed in a zigzag manner. (c) In ms-EPI, multiple acquisitions are performed in an interleaved zigzag manner that collectively span the extent of k-space. The strategy shown here is interleaved EPI.

2.2.1 Distortion in EPI

Despite being the popular choice for DWI, ss-EPI suffers from artifacts that arise due to its strong readout gradients and low bandwidth along the PE direction. These artifacts manifest in different forms, and can be divided into five categories [47]:

1. *Susceptibility Artifacts in the PE Direction*: This artifact is the most problematic in ss-EPI, and arises due to the long readout time in EPI together with inhomogeneities in the B_0 field. Since these inhomogeneities are caused by differences in magnetic susceptibility at tissue-to-tissue and tissue-to-air interfaces [4], the resulting geometric distortions are referred to as susceptibility artifacts. In ss-EPI, the extent of the geometric displacement in the underlying anatomy in the PE direction, $d_{PE}(\mathbf{r})$, is attributed to the echo spacing T_{ESP} of the EPI readout and the field-of-view (FOV) in the PE

direction, FOV_{PE}. Accordingly, $d_{PE}(\mathbf{r})$ can be expressed as [54]

$$d_{PE}(\mathbf{r}) = T_{ESP} FOV_{PE}\ \Delta f(\mathbf{r}), \tag{2.6}$$

where

$$\Delta f(\mathbf{r}) = \frac{\gamma}{2\pi}\ \Delta B_0(\mathbf{r}). \tag{2.7}$$

Here, $\Delta f(\mathbf{r})$ and $\Delta B_0(\mathbf{r})$ denote the off-resonance frequency and B_0 field inhomogeneity, respectively. In addition, $T_{ESP} FOV_{PE}$ is the reciprocal of k-space velocity in the PE direction. The susceptibility artifacts are not only limited to anatomical distortions, but also result in signal pileups and dropouts [5]. An example of this artifact can be seen in Figure 2.4 near the top of the FOV, where severe pileups lead to hyperintensities at the frontal part of the brain. If left uncorrected, these artifacts can result in biased estimates of microstrucural markers, as well as errors during fiber tracking in tractography [6].

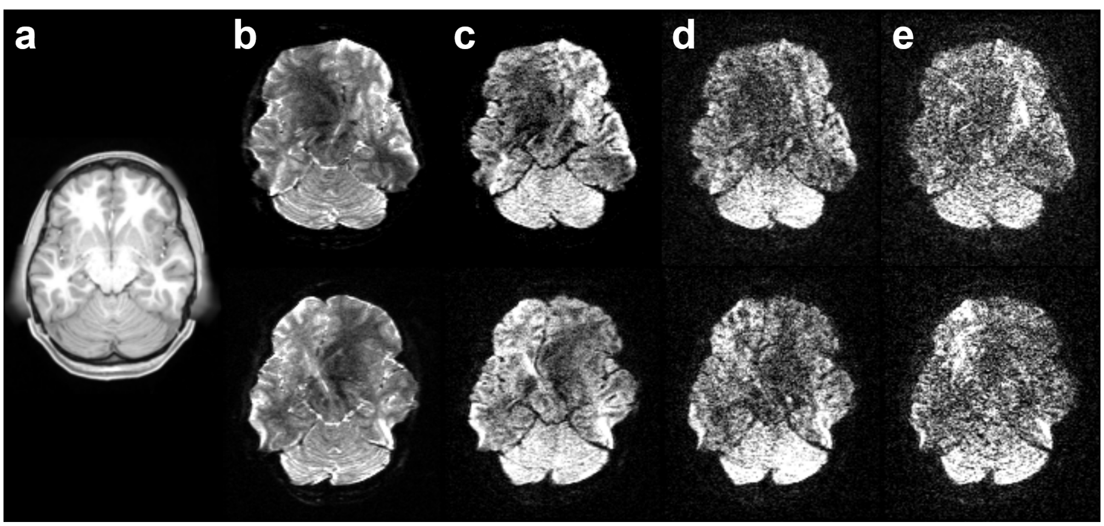

Figure 2.4: Examples of DWI distortion in the brain with different b-values. For each case, the axial view of the same cross-section is shown. (a) a T1 weighted image showing the anatomy. DWI images acquired in reversed PE directions with (b) $b = 0$, (c) $b = 1000\ s/mm^2$, (d) $b = 2000\ s/mm^2$, and (e) $b = 3000\ s/mm^2$. In (b-e), the top row displays the images with left-to-right PE direction and the bottom row displays the images with right-to-left PE direction, showing susceptibility distortions in reversed directions. Data was provided by the Human Connectome Project (see Section 3.1 for more details).

2. *Fat Shift in the PE Direction*: Similar to the susceptibility artifacts discussed above, fat shift artifacts arise from the chemical shift between water and fat. The amount of fat shift is also governed by Equation (2.6), with the off-resonance frequency $\Delta f(\mathbf{r})$ corresponding to the frequency offset stemming from the 3.5 parts per million (ppm) chemical shift of fat with respect to water. Exacerbated by the low bandwidth in the PE direction, these artifacts can be large even at small field strengths, and necessitate some form of fat suppression. If fat is not suppressed, the overlapped fat signal can bias microstrucural markers and lead to complications in the tumor detection process [55, 56].

3. *Eddy Current Induced Geometric Distortion*: This special kind of geometric distortion arises due to rapid switching of gradients, which induce eddy currents in different conducting parts of the scanner [57]. Eddy currents are more severe in diffusion sequences due to the high gradient amplitudes that are employed for the diffusion-sensitizing gradients. These currents act against the gradients that caused them, resulting in a different effective gradient than the one intended. Furthermore, eddy currents due to diffusion-sensitizing gradients can persist for a long duration and affect the data readout, resulting in distortions in the image and spin dephasing that causes a decrease in SNR. Examples of mitigation strategies include using a modified Stejskal-Tanner diffusion-weighting preparation with a single refocusing RF pulse [58], and the twice-refocused spin echo scheme [59]. Eddy current induced distortions affect each EPI readout differently depending on the diffusion-sensitizing direction, and if left uncorrected they can render the acquired data less useful for microstructural analysis and tractography [6].

4. *Nyquist Ghosting*: This artifact is entirely related to the hardware, and occurs due to delays and mismatches during k-space traversal. It manifests as aliasing and shading in the final images [60]. These artifacts can be corrected by applying phase correction based on information from reference scans, by using pulse sequence compensation, or by estimating phase errors from the EPI data itself [61].

5. *Maxwell Effects*: Parabolic shifts occur due to two of Maxwell's equations requiring the curl and divergence of the overall magnetic field to be zero. This issue is systematic, and is compensated for in EPI sequences from major vendors.

As can be concluded from this summary, susceptibility artifacts are the most notorious sources of distortion in EPI. Therefore, correction of these artifacts is imperative for increasing the utility of DWI data acquired with EPI.

2.3 Distortion Correction in EPI

In the category of postprocessing solutions, techniques that rely on reversed PE image pairs to deduce the susceptibility-induced off-resonance field directly from data are predominantly preferred for correction of susceptibility artifacts, as these approaches have been shown to outperform registration-based and measured-field techniques [12]. Correction is commonly performed by using the predicted susceptibility-induced field to directly correct the reversed PE image pairs by unwarping. This unwarping results in blip-up/blip-down corrected images, whose similarity dictates the accuracy of the predicted field. This field can either be represented as a pixel-wise/voxel-wise parametric map that better captures the spatially changing features of the field [13, 14, 10], or as coefficient weights to a predetermined set of basis functions [9].

Popular implementations of this reversed PE method include TOPUP from FSL [15, 9] and HySCO of the SPM toolbox [16, 17]. Further details about TOPUP can be found below. Additionally, this method is highly amenable to the incorporation of deep learning for the purposes of decreasing the time required to predict the susceptibility-induced field and perform correction, as will be discussed soon after. Other correction and mitigation strategies are mentioned lastly, which require additional data or do not completely resolve the problem when compared with reverse PE methods.

2.3.1 Classical Correction Approach

Field-map based susceptibility artifact correction techniques assume that distortions are in opposite directions in the reversed PE image pairs and that only the displacement along the PE direction is significant [10]. In these classical approaches, the displacement field is predicted by performing unwarping correction on the reversed PE image pairs and maximizing their similarity in order to best explain the observed distortions. These reversed PE techniques have been shown to outperform multiple-echo measured-field techniques, and vastly surpassed registration-based techniques in terms of the quality of correction [12].

In this thesis, TOPUP from FSL [15] is taken as the classical "reference" method, as it is often considered a gold standard for EPI distortion correction. TOPUP uses the reversed PE image pairs to estimate the displacement field using unwarping correction [9]. Correction can then be applied by either using least squares restoration to produce the corrected image or by unwarping the reversed PE images and using their average as the corrected image. In TOPUP, unwarping incorporates Jacobian modulation to compensate for intensity pileups that arise from density variations in the field.

The main idea for distortion correction in TOPUP can be described using a matrix representation of the added effect of the acquisition matrix and the transformation between image space and EPI space [9], summarized as follows:

$$\underbrace{\mathbf{f}}_{n_{FE}n_{PE}\times 1} = \underbrace{\mathbf{K}}_{n_{FE}n_{PE}\times n_{FE}n_{PE}} \underbrace{\rho}_{n_{FE}n_{PE}\times 1}, \tag{2.8}$$

where $\mathbf{K}$ is an interpolation matrix called the K-matrix, ρ is the vectorized "true" underlying image, $\mathbf{f}$ is the vectorized EPI image (i.e., the distorted image in EPI space), and n_{PE} and n_{FE} are the image dimensions in the PE and frequency encode (FE) directions, respectively. As shown in Figure 2.5, the K-matrix describes the mapping from the true image to the EPI image. Deviations of the K-matrix from the identity matrix are representative of the amount of distortion, and multiple nonzero values on the same row indicate a many-to-one mapping (i.e., pileup/dropout distortions).

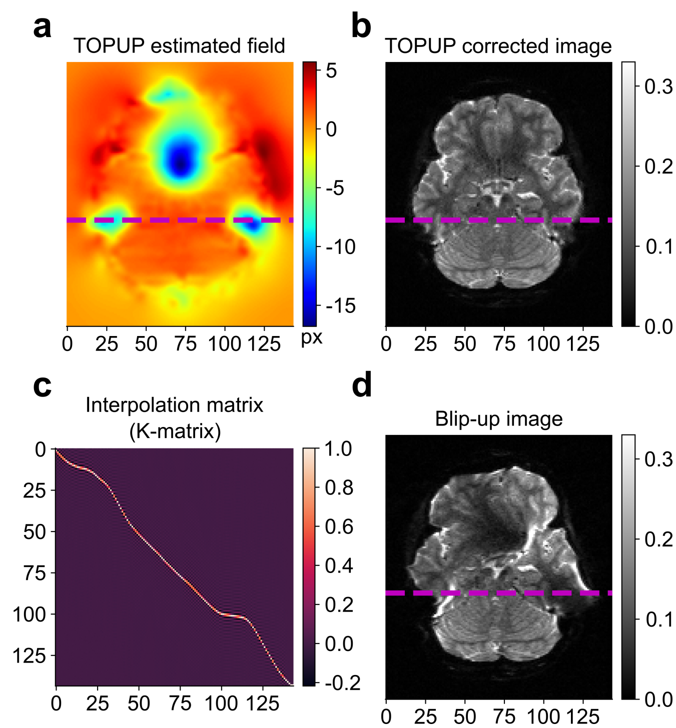

Figure 2.5: Illustration of the image distortion characterized by the K-matrix. (a) The estimated field and (b) the corrected image predicted by TOPUP are shown, with the magenta line highlighting a particular row along the PE direction. (c) The K-matrix formed from the field and (d) the corresponding blip-up image. The deviations of the K-matrix from the identity matrix indicate the amount and direction of distortion, as can be understood by comparing the corrected image and the blip-up image for the highlighted row. The labeled axes correspond to the PE direction.

The K-matrix is large and its implementation poses a problem. Ignoring the distortion along the FE-direction allows block diagonalization and therefore separability of the problem as follows:

$$\mathbf{K} = \begin{bmatrix} \underbrace{\mathbf{K_1}}_{n_{PE}\times n_{PE}} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \underbrace{\mathbf{K_2}}_{n_{PE}\times n_{PE}} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \underbrace{\mathbf{K_{n_{FE}}}}_{n_{PE}\times n_{PE}} \end{bmatrix}, \tag{2.9}$$

Here, $\mathbf{K}_i$, $i = 1, 2, \cdots, n_{FE}$, are the submatrices acting on the true image along the PE-direction, separately for each of the n_{FE} rows.

In general, the K-matrix can be complex-valued for complex-valued images ρ and $\mathbf{f}$, such that it performs a phase shift as well as interpolation [9]. However, the majority of the correction methods, including TOPUP, utilize magnitude images for convenience. Therefore, K-matrix is implemented as a real-valued matrix built from the interpolation model according to the geometry.

TOPUP builds on the K-matrix formulation to predict the field and perform correction, but opts for modelling the field compactly as a linear combination of basis functions [9]. This is in contrast to modelling the field as a voxel-wise parametric map, which allows for increased freedom in capturing spatial changes in the field [13, 14, 10].

2.3.2 Learning-based Correction

Recent deep learning methods in the literature commonly utilize a predicted field to correct the distorted images via unwarping. This field is represented as a pixel- or voxel-wise parametric map. Here, two different learning-based correction methods, S-Net [22] and Deepflow-Net [23], will be covered in detail. These techniques will be used as competing techniques for the proposed FD-Net.

2.3.2.1 S-Net

S-Net [22] utilizes a 3D U-Net [21] to predict the field, followed by unwarping using bilinear interpolation. This unwarping correction module is inspired by the spatial transformer network [62]. For training, S-Net uses local cross-correlation of the corrected blip-up/blip-down images for similarity loss, combined with a diffusion regularizer for field smoothness as follows:

$$\mathcal{L}_{S-Net} = \mathcal{L}_{sim} + \lambda \mathcal{L}_{smooth}, \tag{2.10}$$

where

$$\mathcal{L}_{sim} = 1 - \mathrm{LCC}\left(\mathrm{O}_{unwarp,1}, \mathrm{O}_{unwarp,2}\right). \tag{2.11}$$

Here, λ is the regularization parameter over the smoothness of the displacement field, and $\mathrm{O}_{unwarp,1}$ and $\mathrm{O}_{unwarp,2}$ are the two corrected image outputs of S-Net corresponding to the unwarped reversed PE images. In addition, LCC denotes the local cross-correlation operation, originally proposed as a similarity measure for multimodal image registration [63]. LCC can be expressed as

$$\begin{aligned} \mathrm{LCC}\left(\mathrm{O}_{unwarp,1}, \mathrm{O}_{unwarp,2}\right) &= \sum_{\mathbf{p}\in\Omega} \mathrm{C}_{\mathbf{p}} \\ &= \sum_{\mathbf{p}\in\Omega} \frac{\left(\sum_{\mathbf{p}_i\in \mathrm{W}(\mathbf{p})} \hat{\mathrm{O}}_{unwarp,1}(\mathbf{p}_i)\hat{\mathrm{O}}_{unwarp,2}(\mathbf{p}_i)\right)^2}{\sum_{\mathbf{p}_i\in \mathrm{w}(\mathbf{p})} \left[\hat{\mathrm{O}}_{unwarp,1}(\mathbf{p}_i)\right]^2 \sum_{\mathbf{p}_i\in \mathrm{W}(\mathbf{p})} \left[\hat{\mathrm{O}}_{unwarp,2}(\mathbf{p}_i)\right]^2}, \end{aligned} \tag{2.12}$$

where $\mathrm{C}_{\mathbf{p}}$ is the mean-removed cross-correlation computed voxel-wise for each voxel $\mathbf{p}$ over all voxels $\mathbf{p}_i$ that are within $n \times n \times n$ neighborhood W of $\mathbf{p}$. In addition, $\hat{\mathrm{X}} = \mathrm{X} - \bar{\mathrm{X}}$ denotes the local mean-removed version of X, with $\bar{\mathrm{X}}$ corresponding to the mean of X in the neighborhood W.

In Equation (2.10), $\mathcal{L}_{smooth}$ is defined as the diffusion regularizer of the form

$$\mathcal{L}_{smooth} = \sum_{\mathbf{p}\in\Omega} \left(\frac{\partial \mathrm{O}_{field}(\mathbf{p})}{\partial x}\right)^2 + \left(\frac{\partial \mathrm{O}_{field}(\mathbf{p})}{\partial y}\right)^2 + \left(\frac{\partial \mathrm{O}_{field}(\mathbf{p})}{\partial z}\right)^2, \tag{2.13}$$

where O_{field} is the predicted field output of S-Net. A first-order finite difference is used to approximate the gradients, i.e.,

$$\frac{\partial \mathrm{O}_{field}(\mathbf{p})}{\partial x} \approx \mathrm{O}_{field}\left(\mathrm{p}_x + 1, \mathrm{p}_y, \mathrm{p}_z\right) - \mathrm{O}_{field}\left(\mathrm{p}_x, \mathrm{p}_y, \mathrm{p}_z\right). \tag{2.14}$$

2.3.2.2 Deepflow-Net

Deepflow-Net [23] uses a 2D U-Net architecture and a multiresolution scheme, such that multiscale field predictions from earlier feature stages in the decoder are used in the training, as well. The correction is performed via cubic interpolation, but a density compensation approach similar to that in TOPUP [9] is introduced to compensate for pileups at dense field locations. MSE between the corrected blip-up/blip-down images is used as the similarity measure, and total variation (TV) is used for field smoothness regularization (repeated on all resolution scales), i.e.,

$$\mathcal{L}_{Deepflow-Net} = \sum_m \omega_m \left[\mathcal{L}_{MSE}^{(m)} + \lambda_m \left(\mathcal{L}_{TV}^{(m)} + 10^3 \mathcal{L}_{valley}^{(m)} \right) \right], \tag{2.15}$$

where m refers to the number of multiscale steps, ω_m is the weight of the contribution at each step, and λ_m is the regularization parameter over the smoothness of the field at the corresponding step. In addition, $\mathcal{L}_{MSE}^{(m)}$ is the MSE over each corrected image pair at each resolution scale:

$$\mathcal{L}_{MSE}^{(m)} = \frac{1}{n_{PE}^{(m)} \times n_{FE}^{(m)}} \sum_{\mathbf{p} \in \Omega} \left(\mathrm{O}_{unwarp,1}^{(m)}(\mathbf{p}) - \mathrm{O}_{unwarp,2}^{(m)}(\mathbf{p}) \right)^2, \tag{2.16}$$

where $\mathrm{O}_{unwarp,1}^{(m)}$ and $\mathrm{O}_{unwarp,2}^{(m)}$ are the corrected images at the multiscale step m. Note that care is taken to average over the actual sizes of the images at each multiscale step, $n_{PE}^{(m)} \times n_{FE}^{(m)}$. Next, $\mathcal{L}_{TV}^{(m)}$ is defined as the TV regularizer over the predicted field at each resolution scale, i.e.,

$$\begin{aligned} \mathcal{L}_{TV}^{(m)} = \sum_{(\mathrm{p}_x, \mathrm{p}_y) \in \Omega} & \left| \mathrm{O}_{field}^{(m)} \left(\mathrm{p}_x + 1, \mathrm{p}_y \right) - \mathrm{O}_{field}^{(m)} \left(\mathrm{p}_x, \mathrm{p}_y \right) \right| \\ & + \left| \mathrm{O}_{field}^{(m)} \left(\mathrm{p}_x, \mathrm{p}_y + 1 \right) - \mathrm{O}_{field}^{(m)} \left(\mathrm{p}_x, \mathrm{p}_y \right) \right|. \end{aligned} \tag{2.17}$$

Finally, $\mathcal{L}_{valley}^{(m)}$ is an additional term that prevents the loss function from exploding in the earlier stages of the training, and is given as

$$\mathcal{L}_{valley}^{(m)} = \sum_{\mathbf{p} \in \Omega} \min \left(\left| \mathrm{O}_{field}^{(m)}(\mathbf{p}) \right| - \tau_m, 0 \right), \tag{2.18}$$

where τ_m is a chosen threshold of maximum absolute field swing in units of pixels (chosen as 32 in [23] at full resolution, and scaled appropriately at lower scales). In short, $\mathcal{L}_{valley}^{(m)}$ simply sums the excess amount of the field swing values at occurrences where its magnitude exceeds the threshold. These cases are penalized heaviliy by weighting $\mathcal{L}_{valley}^{(m)}$ with a large constant, such as with 10^3 in Equation (2.15). In later stages of training, $\mathcal{L}_{valley}^{(m}$ can essentially be ignored, once the network converges towards reasonable solutions.

2.3.3 Other Approaches for Distortion Correction

As mentioned above in Subsection 2.3, multiple-echo measured-field techniques and registration-based techniques are also used for correcting susceptibility artifacts. One important drawback of both techniques is the need for additional scans: Multiple-echo measured-field techniques require at least two additional scans that differ only in terms of TE, which are then used to deduce the field from the pixel/voxel phase difference. Registration-based techniques, on the other hand, require additional anatomical reference images to perform registration with the use of a multimodal loss function [20]. Popular implementations are provided in FSL. For example, FMRIB's Utility for Geometrically Unwarping EPIs (FUGUE) employs the multi-echo measured-field approach, while FMRIB's Linear Image Registration Tool (FLIRT) performs image registration based correction [64, 65].

Another class of postprocessing methods relies on the point spread function (PSF) to perform analytic pixel-wise/voxel-wise regularized deconvolution [18]. These methods require auxiliary information such as additional scans and/or further assumptions on the acquired data, such as the k-space trajectory.

Finally, it is worth noting that a class of methods exists that modify the imaging procedure to reduce sensitivity to and alleviate the extent of susceptibility artifacts. Methods such as parallel imaging (e.g., SENSE [66]) decrease sensitivity to the field inhomogeneities by encoding a smaller FOV_{PE} during data acquisition. Other methods attempt to reduce the severity of the susceptibility artifacts by reducing FOV_{PE} directly during excitation, e.g., by using 2D RF pulses to image only the region of interest [67, 68]).

Chapter 3

Methods

This chapter is based in parts on the publication titled:

> A. Zaid Alkilani, T. Çukur, and E. U. Saritas, "A Deep Forward-Distortion Model for Unsupervised Correction of Susceptibility Artifacts in EPI," in *Proceedings of the 30th Annual Meeting of ISMRM*, 2022, p. 0959.

3.1 Experimental Dataset and Setup

For the experiments in this thesis, unprocessed DWI data was randomly selected from Human Connectome Project's (HCP) 1200 Subjects Data Release [69]. The images were acquired on a 3T MRI scanner (Siemens Skyra "Connectom"). A multiband diffusion sequence specific for HCP[1] was utilized, with three different monopolar gradient tables, each acquired using ss-EPI readouts with right-to-left (RL) and left-to-right (LR) reversed PE polarities [70].

[1] A variant was used of the sequence available at `http://www.cmrr.umn.edu/multiband`

The gradient tables include ~90 diffusion-weighting directions, as well as 6 interspersed b0 (i.e., non-diffusion weighted) acquisitions. The diffusion-weighting directions were uniformly distributed over three q-space shells[2] with $b = 1000$, 2000, and 3000 s/mm^2, having roughly equal numbers of acquisitions on each shell. Other imaging parameters include: 210×180 mm^2 field-of-view (FOV), 1.25 mm isotropic resolution, averages = 1, multi-band acceleration factor 3; TR/TE = 5520/89.50 ms, flip angle = 78°; 168×144 acquisition matrix, bandwidth = 1488 Hz/Px, EPI factor = 144, echo spacing = 0.78 μs, and 6/8 phase partial Fourier acquisition.

A total of 20 subjects were selected from the HCP data, 12 for training and 8 for testing. For each subject, a single b0-volume consisting of 111 axial slices with 168×144 image matrix was utilized. TOPUP correction was applied for reference, using the recommended guidelines by the toolbox. The provided configuration file `b02b0_1.cnf` was used for compatibility with the data. The resulting 3D displacement fields and corrected images were sliced axially for compatibility with the network.

All networks were implemented in Keras™ [71] with Tensorflow™ [72] (v2.7.0) backend, on a machine with NVIDIA© GeForce RTX™ 3070 Laptop GPU (GN20-E5 GA104) with 8 GB GDDR6 Graphics Memory and a 256 Bit Memory Bus, Intel® Core™ i7-10875H CPU @2.30GHz with 8 Cores @2304MHz and 16 Logical Processors, and 32GB RAM.

For all networks in this thesis, Adam optimizer was used in training for 1000 epochs until convergence, with early stopping as a form of regularization. Adam optimizer is "computationally efficient, has little memory requirement, invariant to diagonal rescaling of gradients, and is well suited for problems that are large in terms of data/parameters" [73]. A common learning rate of 10^{-4} was empirically set for all experiments, with other parameters set to their default values in Tensorflow.

[2]Using a toolbox from INRIA `https://team.inria.fr/athena/software/`

3.2 Proposed Method: FD-Net

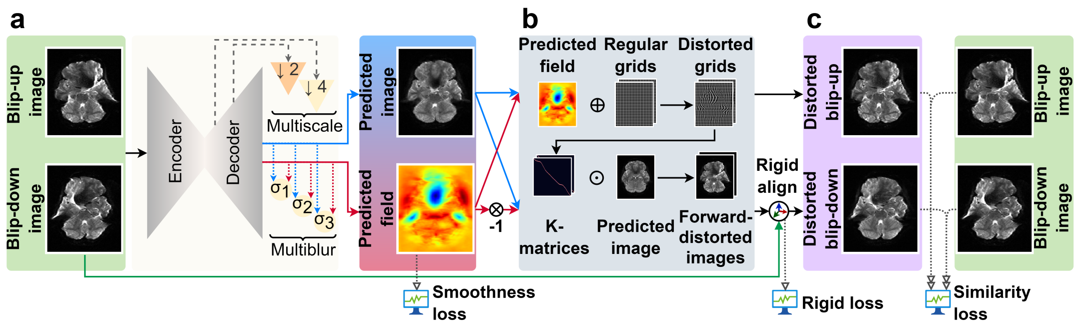

Figure 3.1: Overview of the proposed FD-Net. (a) The input distorted images are fed through an encoder-decoder network, which outputs a predicted field and a predicted image (with optional multiscale and/or multiblur schemes). The field is used to formulate the smoothness loss. (b) The forward-distortion approach of K-Unit is applied in each PE direction, where negating the field is necessary for one of the directions. A rigid alignment is included to improve registration, with the rigid loss formulated from the transformation parameters. (c) The forward-distorted outputs of K-Unit are compared with the input images (redisplayed here for convenience) to formulate the similarity loss. Training is performed over the aggregate of the shown losses.

In this thesis, a novel unsupervised forward-distortion model, FD-Net, is proposed to explicitly constrain measurement fidelity for enhanced correction performance. FD-Net, outlined in Figure 3.1, uses a 2D U-Net to produce both a predicted field and a predicted image from the input reversed PE images. These estimates are then used to enforce consistency to input data together with optional multiresolution schemes: when the predicted image is forward-distorted using the predicted field in both PE directions, the resulting distorted images should match the input reversed PE images (i.e., similarity is enforced in EPI domain). FD-Net outputs a single predicted image, which is motivated by its SNR benefits and is analogous to sensitivity-encoding formulation in parallel imaging that outputs a single combined-coil image [66].

TOPUP has an involved pipeline for implementing and solving the inverse problem (i.e, unwarping correction approach for displacement field prediction). However, unlike TOPUP, FD-Net is only concerned with emulating the distortions. Therefore, matrix inversion is not needed, which removes the majority of the computational complexity of TOPUP (i.e., simple matrix multiplication as in Equation (2.8) will suffice). Moreover, the formulation in Equation (2.9) allows for more efficient computation of the forward-distorted images by reducing the dimensionality of the problem.

Compared to the recent deep learning methods, FD-Net aims to provide a more intuitive approach faithful to the formulation of TOPUP to address the issue of insufficient correction. Note that incorporating physical acquisition parameters (e.g., pixel/voxel resolution in mm) would enable faithful behavior to that of TOPUP. However, doing so would inhibit generalization of performance for deep learning methods across different acquisition conditions, thereby"violating" the premise of using amortized optimization. In contrast, the K-matrix approach in Equation (2.8) can be incorporated without imposing such restrictions on the optimization, while also providing an intuitive yet effective method for directly constraining fidelity to measurements.

3.2.1 Network Architecture

The architecture of the neural networks used in FD-Net are explained in Figure 3.2. In Figure 3.2a, the encoder condenses the information from the input reversed PE images in steps, with a progressively refined receptive field via decreasing filter kernel size and convolutions with stride = 2 for downsampling, to a deep hidden representation. The decoder then attempts to resolve the predicted field and predicted image by building up from this representation, by means of upsampling and concatenation of the encoder steps using skip connections. These skip connections facilitate gradient and information flow which helps ease training and avoid the vanishing gradient problem. All convolutional layers use Leaky Rectified Linear Unit (ReLU) activation with a slope coefficient $\alpha = 0.2$, except at the final steps of the decoder as indicated in the Figure 3.2a; the predicted image is formed from a convolutional layer with ReLU activation and the predicted field is formed using a linear activation.

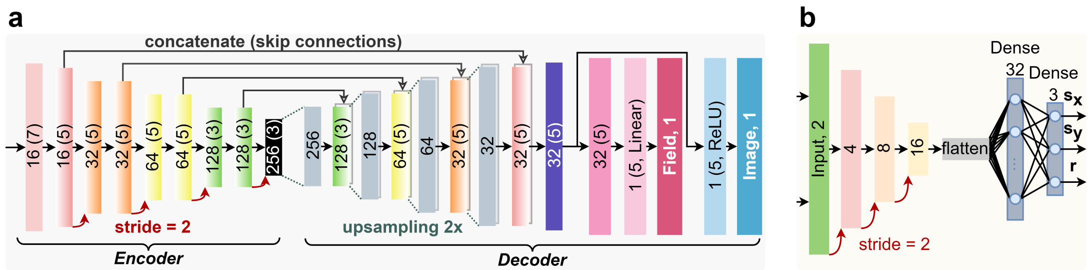

Figure 3.2: Details of the neural networks employed in FD-Net. (a) The encoder-decoder network is outlined with blocks representing the convolutional steps. Convolution with stride 2 is used to reduce dimensionality in the encoder steps, while upsampling by 2 is used to increase it in the decoder steps. Skip connections are introduced to facilitate information flow and improve gradient propagation by concatenating the encoder representations to the corresponding stages in the decoder. The numbers inside the boxes denote the feature dimensions, with the numbers in brackets indicating the filter kernel sizes. Leaky ReLU activation with slope coefficient of 0.2 is used, unless otherwise indicated. (b) A rigid transformation network is used to align one of the forward-distorted images to its corresponding input PE image. The first stage encodes the images using convolutional layers with stride of 2, shown as boxes with the numbers inside indicating the feature dimensions. The output of the convolutional stage is flattened and passed through a dense layer with 32 neurons and another with 3 neurons. The final output comprises the three parameters required for the rigid transformation to be applied to the forwarded-distorted image.

In Figure 3.2b, the rigid transformation network[3] accepts two forward-distorted images as inputs, condenses them using convolutions and then flattens the representation to plug into dense layers that transform the output to three numbers s_x, s_y, and r, which correspond to the transformation parameters describing the x-axis shift, y-axis shift, and in-plane rotation, respectively. These parameters are then used to apply a rigid transformation to only one of the forward-distorted images (blip-down distorted image in this case) to improve alignment with the corresponding input distorted image. Note that a similar rigid alignment is also performed in TOPUP, and it offloads some burden from the non-rigid field-based alignment process by accounting for non-idealities such as subject movements between the two reversed PE scans.

[3]Adapted from `https://github.com/oarriaga/STN.keras.git` with a few changes, including using bicubic instead of bilinear interpolation to boost performance.

3.2.2 K-Unit

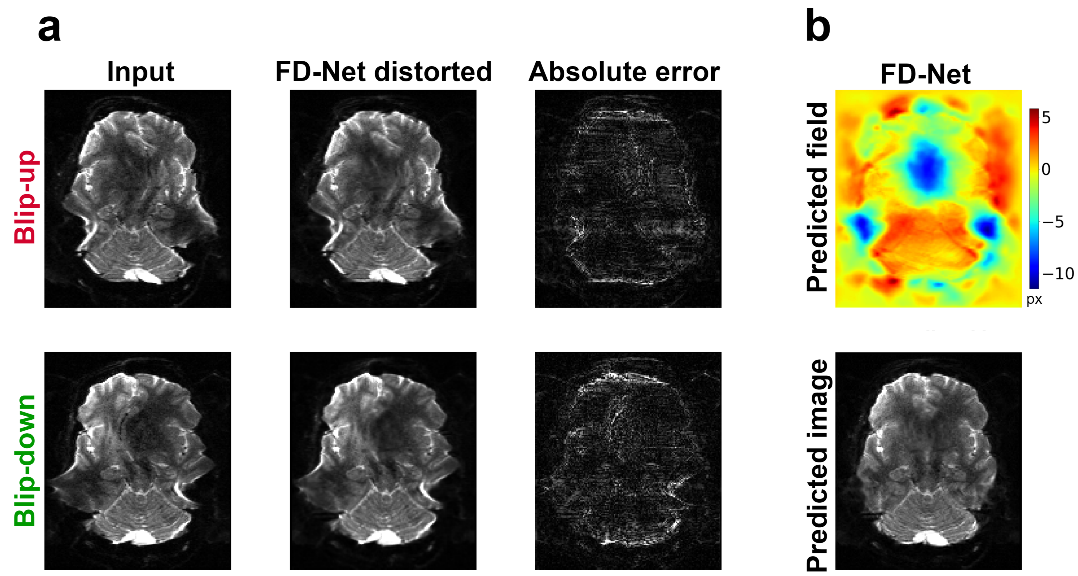

Figure 3.3: Example of forward distortion by using the K-Unit in FD-Net. (a) The input blip-up and blip-down EPI images are compared with the results of forward distortion in FD-Net, with the absolute error maps provided in the far right column. The display window for the error is exaggerated in order to view the details of the error. (b) The predicted field and predicted image results that are used in the K-Unit for forward distortion.

The K-Unit in FD-Net performs forward distortion on the estimated image using the estimated field. The columns of the K-matrix define the neighborhood of pixels/voxels, and therefore can be generated using any valid interpolation kernel. For the 1-D case, the interpolation operation can be expressed as

$$u(x) = \sum_{i=1}^{N} \kappa(x - x_i) v(x_i), \tag{3.1}$$

where $u(x)$ is the interpolant, $v(x)$ denotes the function/values to be interpolated, and κ is the interpolation kernel. Here, we utilize a sinc kernel when constructing the K-matrix, i.e.,

$$\kappa(\xi) = \text{sinc}(\xi) = \frac{\sin(\pi\xi)}{\pi\xi}. \tag{3.2}$$

The K-matrix approach for the 2D case can be applied via a series of K-submatrices. As described in Equation (2.9), this approach exploits the separability of the problem and is therefore computationally efficient. Building on this separability, the first step is the formation of a uniform grid $\mathbf{X}_{grid}$:

$$\underbrace{\mathbf{X}_{grid}}_{n_{PE}\times n_{PE}} = \begin{bmatrix} 1 & \cdots & 1 \\ 2 & \cdots & 2 \\ \vdots & & \vdots \\ n_{PE} & \cdots & n_{PE} \end{bmatrix}. \tag{3.3}$$

The distorted grid after the interpolation, $[\mathbf{X}_{interp}]_i$ is formed by determining the new grid location for each pixel from the shift amount given in the displacement field, i.e.,

$$\underbrace{[\mathbf{X}_{interp}]_i}_{n_{PE}\times n_{PE}} = \begin{bmatrix} \mathrm{O}_{field}(i,1)+1 & \cdots & \mathrm{O}_{field}(i,1)+1 \\ \mathrm{O}_{field}(i,2)+2 & \cdots & \mathrm{O}_{field}(i,2)+2 \\ \vdots & & \vdots \\ \mathrm{O}_{field}(i,n_{PE})+n_{PE} & \cdots & \mathrm{O}_{field}(i,n_{PE})+n_{PE} \end{bmatrix}, \tag{3.4}$$

where O_{field} is the estimated field output of FD-Net and $i = 1, 2, \ldots, n_{FE}$ is the row index over the FE direction. For practical purposes, each entry in $[\mathbf{X}_{interp}]_i$ is kept limited between 1 and n_{PE} (i.e., clipped to the valid range of interpolation). Note that for the other PE direction, the negative of the displacement field is utilized instead.

Taking the difference of the two grids and then applying the kernel gives us the K-matrix that will act on the i^{th} row as follows:

$$\underbrace{[\mathbf{K}]_i}_{n_{PE}\times n_{PE}} = \kappa\left([\mathbf{X}_{interp}]_i - \mathbf{X}_{grid}\right). \tag{3.5}$$

Using this K-matrix, the i^{th} row of the forward-distorted image is formed with a matrix multiplication:

$$\underbrace{\left[\mathrm{O}^T_{dist}\right]_i}_{n_{PE}\times 1} = [\mathbf{K}]_i \underbrace{\left[\mathrm{O}^T_{image}\right]_i}_{n_{PE}\times 1}, \tag{3.6}$$

where $(\cdot)^T$ denotes the matrix transpose, O_{image} denotes the predicted "true" image, and O_{dist} denotes the forward-distorted image.

Finally, the full forward-distorted image O_{dist} can be formed by stacking $\left[\mathrm{O}_{dist}^{T}\right]_i$ for all i, i.e.,

$$\underbrace{\mathrm{O}_{dist}}_{n_{FE}\times n_{PE}} = \left[\left[\mathrm{O}_{dist}^{T}\right]_1 \;\middle|\; \left[\mathrm{O}_{dist}^{T}\right]_2 \;\;\cdots \;\middle|\; \left[\mathrm{O}_{dist}^{T}\right]_{n_{FE}}\right]^{T}. \tag{3.7}$$

The rows of the K-matrix perform intensity modulation, where the matrix multiplication with the "true" image in Equation (3.6) results in "modulated interpolation" using the pixel/voxel neighbors and leads to signal pileups/dropouts, as intended. This process is demonstrated in Figure 3.3, which provides an illustrative example of forward distortion in FD-Net using the K-Unit.

3.2.3 Loss

The loss for FD-Net is described as

$$\mathcal{L}_{FD-Net} = \sum_{m} \omega_m \left[\mathcal{L}_{MSE}^{(m)} + \lambda_m \left(\mathcal{L}_{BE}^{(m)} + 10^3 \mathcal{L}_{valley}^{(m)}\right)\right] + \gamma \mathcal{L}_{rigid}, \tag{3.8}$$

where m is the index of multiresolution step, and ω_m and λ_m are the weighting parameter and the regularization parameter over the smoothness of the field for the multiresolution step m, respectively. In addition, γ is the contribution of the rigid loss. Here, the first term is the summation of the forward-distortion loss contributions, while the second term is the rigid loss contribution. Each loss term is described in detail below.

First, $\mathcal{L}_{MSE}^{(m)}$ is the MSE over the input reversed PE and the forward-distorted image pairs at each resolution level:

$$\begin{aligned}\mathcal{L}_{MSE}^{(m)} = \frac{1}{2 \times n_{PE}^{(m)} \times n_{FE}^{(m)}} \Bigg[& \sum_{\mathbf{p}\in\Omega} \left(\mathrm{O}_{dist,1}^{(m)}(\mathbf{p}) - I_{im,1}^{(m)}(\mathbf{p})\right)^2 \\ & + \sum_{\mathbf{p}\in\Omega} \left(\mathbf{O}_{dist,2}^{(m)}(\mathbf{p}) - I_{im,2}^{(m)}(\mathbf{p})\right)^2 \Bigg],\end{aligned} \tag{3.9}$$

where $I_{im,1}$ and $I_{im,2}$ denote the input reversed PE images. Note that care is taken to average over the actual size of the images at each multiresolution step, $n_{PE}^{(m)} \times n_{FE}^{(m)}$. Division by 2 accounts for the two PE directions.

Next, $\mathcal{L}_{BE}^{(m)}$ is defined as the bending energy regularizer [74] over the field at each multiresolution step, which can be expressed as

$$\mathcal{L}_{BE}^{(m)} = \sum_{\mathbf{p}\in\Omega} \left(\frac{\partial^2 \mathrm{O}_{field}^{(m)}(\mathbf{p})}{\partial x^2}\right)^2 + \left(\frac{\partial^2 \mathrm{O}_{field}^{(m)}(\mathbf{p})}{\partial y^2}\right)^2 + \left(\frac{\partial^2 \mathrm{O}_{field}^{(m)}(\mathbf{p})}{\partial xy}\right)^2 + \left(\frac{\partial^2 \mathrm{O}_{field}^{(m)}(\mathbf{p})}{\partial yx}\right)^2 . \tag{3.10}$$

In practice, finite differences [75] are used to approximate the gradients. First-order finite difference is used twice to approximate the second-order derivative cross-terms, e.g.,

$$\frac{\partial \mathrm{O}_{field}(\mathbf{p})}{\partial x} \approx -0.5\mathrm{O}_{field}\left(p_x - 1, p_y, p_z\right) + 0.5\mathrm{O}_{field}\left(p_x + 1, p_y, p_z\right). \tag{3.11}$$

Likewise, second-order finite difference is used to approximate the second-order derivative direct-terms, e.g.,

$$\frac{\partial^2 \mathrm{O}_{field}^{(m)}(\mathbf{p})}{\partial x^2} \approx \mathrm{O}_{field}\left(p_x - 1, p_y, p_z\right) - 2\mathrm{O}_{field}\left(p_x, p_y, p_z\right) + \mathrm{O}_{field}\left(p_x + 1, p_y, p_z\right). \tag{3.12}$$

As mentioned in Subsection 3.2.1, rigid alignment of the forward-distorted image and the input EPI image for one of the PE directions is achieved via the subnetwork illustrated in Figure 3.2b. The loss term $\mathcal{L}_{rigid}$ aims to find the smallest possible rigid transformation parameters that would perform this alignment, which would ideally be close to all zeroes. Thus, $\mathcal{L}_{rigid}$ is defined as follows:

$$\mathcal{L}_{rigid} = s_x^2 + s_y^2 + r^2. \tag{3.13}$$

Because the same rigid alignment parameters apply to all multiresolution levels, the rigid loss is only included once in Equation (3.8).

3.2.4 Multiresolution Scheme

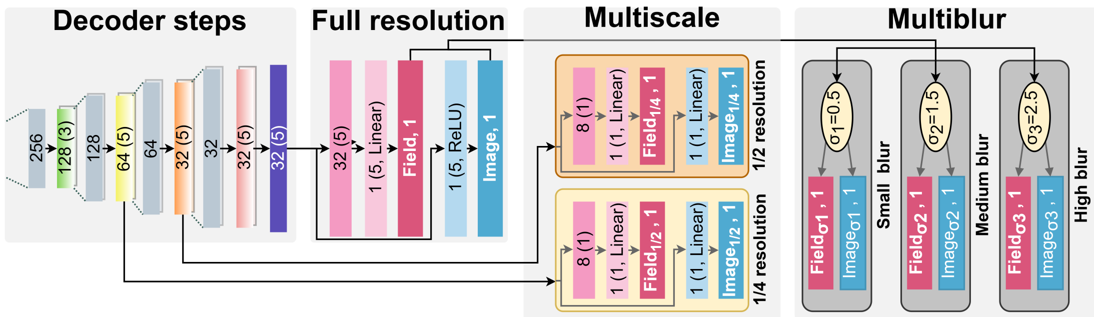

Figure 3.4: Illustration of the two multiresolution approaches, multiscale and multiblur, considered for FD-Net. The last stages of the decoder that generate the full resolution image and field are indicated for clarity. The multiscale approach relies on forming an output image and field at a lower dimensional scale, using appropriate convolutional steps to produce the outputs at 1/2 and 1/4 scale in this case. The numbers inside the boxes denote the feature dimensions, with the numbers in brackets indicating the filter kernel sizes. Leaky ReLU activation with slope coefficient of 0.2 is used, unless otherwise indicated. For the multiblur case, Gaussian blur kernels are applied to the full resolution outputs to create increasingly blurred results. Blur amounts of $\sigma_1 = 0.5$, $\sigma_2 = 1.5$, and $\sigma_3 = 2.5$ are used in this case, with Gaussian kernel size of $\lceil 4\sigma_m \rceil$ in each case. The results of all the incorporated multiresolution levels are passed through the K-Unit and used to contribute to the overall loss in a regularizing manner.

Multiresolution schemes aim to improve performance by enforcing consistency across different scales and blurs, in principle leading to faster convergence and less chance of encountering local minima. Similar to TOPUP, FD-Net supports multiresolution schemes, where the output is evaluated at different scales of the resolution, which we refer to as multiscale, and/or at different blur amounts, which we refer to as multiblur. Figure 3.4 illustrates the modifications and additions necessary to accomplish these multiresolution schemes.

For the multiscale case, different stages of the decoder are used to predict the field and image at lower resolutions of the predicted outputs. To accomplish this, convolutional layers akin to the full resolution case are employed to form the predicted field and image, at 1/2 and 1/4 of the full scale in this case. These predictions are fed through the same K-Unit structure as the full resolution case, after appropriately scaling their contribution to the overall loss, including scaling up their smoothness parameter λ_m in Equation (3.8) as well as their corresponding τ_m (Equation (2.18)).

As for the multiblur case, the full resolution output is blurred with a Gaussian kernel of increasing blur amounts σ_m with kernel size $\lceil 4\sigma_m \rceil$. These blurred predictions are also fed through the same K-Unit structure as the full resolution case, after appropriately scaling their contribution to the overall loss by scaling up their smoothness parameter λ_m in Equation (3.8).

For both multiresolution strategies, the input PE images are also processed similarly to enable computation of the loss term $\mathcal{L}_{MSE}^{(m)}$. For the multiscale case, the input PE images were resized and appropriately blurred to avoid aliasing artifacts. For the multiblur case, only blurring is necessary. A similar treatment can be applied to TOPUP results whenever needed to facilitate comparison, but is not necessary when only the full resolution results are to be considered.

The performances of the multiscale and multiblur schemes, as well their combination, were compared to determine the best multiresolution strategy. The hyperparameters chosen for each multiresolution scheme considered are provided in Table 3.1. As will be demonstrated in Chapter 4, the multiblur case stands out in terms of performance. Therefore, an ablation study for the multiblur strategy was performed with different combinations of three different blur amounts, as summarized in Table 3.2. Here, S, M, and H refer to small, medium, and high blur cases, respectively. The parameters were chosen empirically, with the purpose of maximizing quantitative image quality metrics with respect to TOPUP over the predicted field and the predicted image (see Section 3.3 below).

Table 3.1: Hyperparameter choices for the multiresolution strategies in FD-Net. The top row lists the hyperparameters, with ω split into its constituent full resolution ("FR"), multiscale ($\frac{1}{2}$ and $\frac{1}{4}$ scale), and multiblur ($\sigma_1 = 0.5$, $\sigma_2 = 1.5$, and $\sigma_3 = 2.5$ blur) components. For each multiresolution strategy, irrelevant hyperparameters are disregarded and marked with a dash (−). The hyperparameters considered are: the field regularization weight λ, which defines the contribution of field smoothness; the multiresolution weight ω, which defines the contribution of the forward-distortion loss; the rigid loss weight γ, which defines the contribution of the loss over the rigid alignment network transformation parameters; the valley loss threshold τ, which defines the maximum permissible field swing.

Proposed FD-Net	λ	ω						γ	τ
		FR	$\frac{1}{2}$	$\frac{1}{4}$	σ_1	σ_2	σ_3		
No multiresolution	10^{-5}	1.0	−	−	−	−	−	0.01	32
Multiscale		0.6	0.3	0.1	−	−	−		
Multiblur		0.4	−	−	0.3	0.2	0.1		
Multiscal & multiblur		0.5	0.15	0.05	0.15	0.1	0.05		

Table 3.2: Hyperparameter choices for the multiblur scheme ablation study for FD-Net. The top row lists the hyperparameters, with ω split into its constituent full resolution ("FR") and multiblur ($\sigma_1 = 0.5$, $\sigma_2 = 1.5$, and $\sigma_3 = 2.5$ blur) components. The leftmost column denotes the combinations over the considered three blur amounts for the multiblur case, namely: small (S) for the σ_1 blur, Medium (M) for the σ_2 blur, and high (H) for the σ_3 blur. For each combination, irrelevant blurs are disregarded and marked with a dash (−). The hyperparameters considered are: the field regularization weight λ, which defines the contribution of field smoothness; the multiresolution weight ω, which defines the contribution of the forward-distortion loss; the rigid loss weight γ, which defines the contribution of the loss over the rigid alignment network transformation parameters; the valley loss threshold τ, which defines the maximum permissible field swing.

Proposed FD-Net	λ	ω_m				γ	τ
		FR	σ_1	σ_2	σ_3		
S-M-H	10^{-5}	0.4	0.3	0.2	0.1	0.01	32
M-H		0.57	−	0.29	0.14		
S-H		0.5	0.38	−	0.13		
S-M		0.45	0.33	0.22	−		
H		0.8	−	−	0.2		
M		0.67	−	0.33	−		
S		0.57	0.47	−	−		

3.3 Competing Methods

Two unsupervised learning-based methods, S-Net and Deepflow-Net, were implemented for comparison. In addition, a supervised network was implemented to serve as a baseline for FD-Net. Implementations of competing methods were maintained as faithful to FD-Net as possible to facilitate comparison:

1. *S-Net:* As described in Subsection 2.3.2, S-Net was implemented with correction performed using a modified K-Unit approach. The approach is essentially similar to the one applied by Duong *et al.* [22], and can be achieved with minimal difference to Equation (3.6). This is possible by exploiting the nature of the K-matrix:

   $$\underbrace{\left[\mathrm{O}_{unwarp}^{T}\right]_i}_{n_{PE}\times 1} = [\mathbf{K}]_i^T \underbrace{\left[\mathrm{I}_{PE}^{T}\right]_i}_{n_{PE}\times 1}, \tag{3.14}$$

   where I_{PE} denotes the input distorted PE under consideration and O_{unwarp} denotes the corresponding unwarped image. By transposing $[\mathbf{K}]_i$, intensity modulation is forgone in favor of performing standard interpolation. As such, only the field head at the end of the decoder in Figure 3.2a was necessary. Unwarping of the two input PE images was performed separately, and the average of the two resulting images was taken as the corrected image. Implementation of LCC was adapted from the GitHub repo[4] of Voxelmorph [76]. In place of the diffusion regularizer from Equation (2.13), bending energy from Equation (3.10) was used in order to facilitate comparison with FD-Net. Similarly, the rigid transformation network was utilized and the rigid loss from Equation (3.13) was incorporated into the training. No density compensation was incorporated, as is done by Duong *et al.* [22].

2. *Deepflow-Net:* Deepflow-Net was implemented with density compensated unwarping correction based on a modified K-Unit approach. The K-matrix can be multiplied with an image of all 1's, **1**, to produce a density pileup

[4]Available at `https://github.com/voxelmorph/voxelmorph`

map $\mathbf{W}$, which can be inverted. When used to weight the input PE image, this process essentially emulates the density compensation approach of Zahneisen *et al.* [23]:

$$\underbrace{\left[\mathrm{O}^{T}_{unwarp}\right]_i}_{n_{PE}\times 1} = [\mathbf{K}]_i^T \underbrace{\left((\mathbf{1} \oslash \mathbf{W}) \odot \left[\mathrm{I}^{T}_{PE}\right]_i\right)}_{n_{PE}\times 1}, \tag{3.15}$$

where

$$\underbrace{\mathbf{W}}_{n_{PE}\times 1} = [\mathbf{K}]_i \underbrace{\mathbf{1}}_{n_{PE}\times 1}. \tag{3.16}$$

Here, $\oslash$ and $\odot$ denote Hadamard division and product, respectively, and $\mathbf{W}$ is limited in $[0, 1]$. Similar to Equation (3.14), this approach also exploits the nature of the K-matrix and enables unwarping correction in an intuitive manner. In the same manner as S-Net, only the field was generated and the average of the two unwarped images was used as the corrected image. The multiscale strategy was applied in a similar manner to that of FD-Net, and TV regularization from Equation (2.17) was replaced with bending energy from Equation (3.10). The rigid transformation network was also incorporated along with its loss term. These modifications and additions to what was described in Subsection 2.3.2 permit a fair comparison with FD-Net.

3. *Supervised Baseline:* Finally, a supervised baseline was trained with an architecture identical to that of FD-Net, with the exception of the loss being completely supervised. For this purpose, the network predicted field and predicted image were matched with the results from TOPUP as follows:

$$\begin{aligned}\mathcal{L}_{Supervised} = \frac{1}{2 \times n_{PE} \times n_{FE}} \Bigg[&\sum_{\mathbf{p}\in\Omega} \left(\mathrm{O}_{image}(\mathbf{p}) - \mathrm{T}_{image}(\mathbf{p})\right)^2 \\ &+ \sum_{\mathbf{p}\in\Omega} \left(\mathrm{O}_{field}(\mathbf{p}) - \mathrm{T}_{field}(\mathbf{p})\right)^2 \Bigg].\end{aligned} \tag{3.17}$$

Here, T_{image} and T_{field} denote the corrected image and field provided from TOPUP, respectively.

3.4 Quantitative Assessments

In order to permit quantitative image quality assessments (IQA) for FD-Net and the above-mentioned competing methods, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) were utilized. TOPUP results were taken as the reference. Accordingly, both the predicted field and predicted image were compared with the TOPUP results. Before computing PSNR and SSIM, the field generated by each method was first masked via a median Otsu threshold over the TOPUP image to remove background regions from consideration.

The hyperparameters chosen for each of the competing methods and FD-Net are provided in Table 3.3. For the competing methods, the parameters were chosen empirically, to maximize PSNR and SSIM with respect to TOPUP results for each method, individually.

Table 3.3: Hyperparameter choices for competing methods and the proposed FD-Net. The top row lists the hyperparameters, with ω split into its constituent full resolution ("FR") and multiscale ($\frac{1}{2}$ and $\frac{1}{4}$ scale) components. The leftmost column denotes each of the competing methods considered. For each method, irrelevant hyperparameters are disregarded and marked with a dash ($-$). The hyperparameters considered are: the field regularization weight λ, which defines the contribution of field smoothness; the multiresolution weight ω, which defines the contribution of the forward-distortion loss; the rigid loss weight γ, which defines the contribution of the loss over the rigid alignment network transformation parameters; the valley loss threshold τ, which defines the maximum permissible field swing.

<table>
<tr><th rowspan="2">Methods</th><th rowspan="2">λ</th><th colspan="6">ω_m</th><th rowspan="2">γ</th><th rowspan="2">τ</th></tr>
<tr><th>FR</th><th>$\frac{1}{2}$</th><th>$\frac{1}{4}$</th><th>σ_1</th><th>σ_2</th><th>σ_3</th></tr>
<tr><td>Proposed FD-Net</td><td rowspan="2">10^{-5}</td><td>0.4</td><td>$-$</td><td>$-$</td><td>0.3</td><td>0.2</td><td>0.1</td><td rowspan="3">0.01</td><td rowspan="2">32</td></tr>
<tr><td>Deepflow-Net</td><td>0.6</td><td>0.3</td><td>0.1</td><td>$-$</td><td>$-$</td><td>$-$</td></tr>
<tr><td>S-Net</td><td>10</td><td>1.0</td><td>$-$</td><td>$-$</td><td>$-$</td><td>$-$</td><td>$-$</td><td>$-$</td></tr>
<tr><td>Supervised baseline</td><td>$-$</td><td>1.0</td><td>$-$</td><td>$-$</td><td>$-$</td><td>$-$</td><td>$-$</td><td>$-$</td><td>$-$</td></tr>
</table>

Chapter 4

Results

This chapter presents the results of training and evaluating the performance of each method with respect to TOPUP. The choice of multiresolution strategy for FD-Net is first justified by means of quantitative analyses, and the particular choice of the blur combination is also detailed in a separate ablation study. Afterwards, evaluations of the competing methods in terms of both quantitative and visual results are performed, in order to establish the merit of the proposed FD-Net in terms of correction/prediction performance.

As a general note on runtime, correction of a volume took on average ~7.5 seconds for each network considered. TOPUP took on average ~3086 seconds (~51.5 minutes) to predict the field and an additional ~6 seconds to apply correction.

4.1 Multiresolution Schemes for FD-Net

We first evaluate each multiresolution strategy for the proposed FD-Net, alongside a strategy with no multiresolution. Using the hyperparameters from Table 3.1, FD-Net was trained and subsequently evaluated for each strategy. The results are summarized in Table 4.1 and Table 4.2 for PSNR and SSIM, respectively.

Table 4.1: PSNR performance of the multiresolution strategies in FD-Net. The leftmost column indicates the multiresolution strategy considered, and the rightmost two columns indicate the PSNR performance of the predicted image quality and the field quality, respectively. PSNR is measured in dB with respect to the corresponding TOPUP results. The mean and standard deviation of PSNR across the subjects are reported. The best performing result in each column is highlighted in bold font. The multiblur strategy is chosen as the multiresolution scheme for FD-Net due to its superior performance.

Proposed FD-Net	Image quality	Field quality
No multiresolution	30.62 (2.82)	20.80 (6.11)
Multiscale	30.31 (2.67)	20.00 (6.14)
Multiblur	**31.21 (2.76)**	**22.30 (5.71)**
Multiscale&multiblur	30.88 (2.78)	21.67 (5.82)

Table 4.2: SSIM performance of the multiresolution strategies in FD-Net. The leftmost column indicates the multiresolution strategy considered, and the rightmost two columns indicate the SSIM performance of the predicted image and the field, respectively. SSIM is provided as a percentage, measured with respect to the corresponding TOPUP results. The mean and standard deviation of SSIM across the subjects are reported. The best performing result in each column is highlighted in bold font. The multiblur strategy is chosen as the multiresolution scheme for FD-Net due to its superior performance.

Proposed FD-Net	Image quality	Field quality
No multiresolution	84.60 (11.72)	80.15 (11.85)
Multiscale	83.91 (11.72)	79.24 (12.44)
Multiblur	**86.49 (11.53)**	**82.96 (10.38)**
Multiscale&multiblur	85.84 (11.79)	82.52 (10.42)

It is immediately evident that introducing a multiblur strategy provides a performance boost, and that the multiscale strategy underperforms in comparison to both the multiblur and the no multiresolution cases. Using the multiblur strategy, the image quality is improved by 0.59dB PSNR/1.89% SSIM, and the field quality is improved by 1.50dB PSNR/2.81% SSIM over the no multiresolution case. A combination of multiblur and multiscale strategies does not improve over the multiblur case, indicating that multiblur alone is sufficient to boost performance.

4.1.1 Multiblur Ablation Study for FD-Net

To further investigate the efficacy of the multiblur strategy, an ablation study was conducted with different choices of blur amounts. Using the hyperparameters in Table 3.2, different combinations of three different blur amounts were considered and the performance of FD-NET was evaluated for each case. The results are provided in Table 4.3 and Table 4.4 for PSNR and SSIM, respectively.

The results show that including two or more blur stages boosts performance, with the M-H and S-M-H multiblur combinations providing the best results. While these two multiblur combinations show similar performances, the S-M-H multiblur combination is chosen as the multiresolution strategy for FD-Net due to its high performance, with the potential to provide a reliable generalization by incorporating all blur stages. The S-M-H multiblur combination improves the image quality by 2.20dB PSNR/5.01% SSIM and the field quality by 5.81dB PSNR/∼8.67% SSIM for field quality over the worst performing multiblur scheme.

Table 4.3: PSNR performance for the multiblur scheme ablation study for FD-Net. The leftmost column indicates the combination of the three blur amounts considered for the multiresolution strategy: small (S) , medium (M), and high (H) blur. The rightmost two columns indicate PSNR performance of the predicted image and the field, respectively. PSNR is measured in dB with respect to the corresponding TOPUP results. The mean and standard deviation of PSNR across the subjects are reported. The best performing result in each column is highlighted in bold font. The S-M-H multiblur combination is chosen as the multiresolution scheme for FD-Net due to its competitive PSNR performance.

Proposed FD-Net	Image quality	Field quality
S-M-H	31.21 (2.76)	22.30 (5.71)
M-H	**31.26 (2.87)**	**22.39 (5.61)**
S-H	31.03 (2.79)	21.60 (5.91)
S-M	31.01 (2.73)	21.78 (5.84)
H	30.96 (2.87)	21.52 (5.75)
M	29.01 (2.84)	16.49 (6.94)
S	30.73 (2.68)	21.12 (5.99)

Table 4.4: SSIM performance for the multiblur scheme ablation study for FD-Net. The leftmost column indicates the combination of the three blur amounts considered for the multiresolution strategy: small (S) for the $\sigma_1 = 0.5$ blur, Medium (M) for the $\sigma_2 = 1.5$ blur, and high (H) for the $\sigma_3 = 2.5$ blur. The rightmost two columns indicate SSIM performance of the predicted image and the field, respectively. SSIM is provided as a percentage, measured with respect to the corresponding TOPUP results. The mean and standard deviation of SSIM across the subjects are reported. The best performing result in each column is highlighted in bold font. The S-M-H multiblur combination is chosen as the multiresolution scheme for FD-Net due to its superior SSIM performance.

Proposed FD-Net	Image quality	Field quality
S-M-H	**86.49 (11.53)**	**82.96** (10.38)
M-H	86.44 (11.66)	82.79 (10.52)
S-H	85.98 (11.75)	81.94 (10.94)
S-M	85.64 (11.61)	82.14 (10.84)
H	85.83 (11.73)	81.51 (11.19)
M	81.48 (11.74)	74.29 (14.91)
S	84.80 (11.56)	80.80 (11.60)

4.2 Comparison with Competing Methods

As mentioned earlier in Chapter 3, the competing methods from literature were implemented as faithfully as possible, while facilitating comparison with the proposed FD-Net (see Table 3.3 for a summary of the hyperparameters). The supervised FD-Net implementation serves as a baseline for the unsupervised methods, particularly to demonstrate the success of FD-Net as an unsupervised method. The quantitative and visual results provided below aim to establish this claim.

4.2.1 Quantitative Results

A comprehensive quantitative evaluation between the proposed FD-Net and the competing methods was conducted by means of PSNR and SSIM computed with respect to the TOPUP results. Slice-wise and subject-wise results are demonstrated graphically, in addition to the tabulated results provided for numerical comparison:

1. *Slice-Wise Evaluation:* To demonstrate the performances across different slices of the dataset, Figure 4.1b was generated showing the 95% confidence interval around the mean, for PSNR and SSIM of both the predicted image and field from each method. Since the acquisitions were conducted to capture the same anatomy at the same orientation for all subjects, a given slice number captured approximately the same anatomy in all subjects. Therefore, no additional intersubject registration was conducted for this analysis. To illustrate the underlying anatomy, a representative T1 weighted image from one of the subjects is provided in Figure 4.1a, with slices that will be relevant in Subsection 4.2.2 highlighted with a magenta dashed lines. For each subject, registration of the T1 weighted volume to the b0 volume was performed using FSL's FLIRT [64, 65]. The results in Figure 4.1b show that all methods have dips in performance at the same slice indices, providing insight into which slices are challenging in terms of distortion correction. Likewise, the peaks in performances occur at similar slices indices, corresponding to distortions that are not as severe, i.e., a less challenging problem for the methods to solve. FD-Net outperforms all competing methods in terms of the predicted image quality, especially at the problematic lower brain slices where large distortions are present. Moreover, the predicted field quality from FD-Net exceeds the competing methods, except for the supervised baseline.

2. *Subject-Wise Evaluation:* The performance of each method was also assessed by computing the mean PSNR and SSIM for both the predicted image and field, over all slices in the volume of a given subject, for each of the 8 subjects reserved for testing. The results are plotted as a scatter of FD-Net vs. each competing method as shown in Figure 4.2, with results above the dashed identity line representing superior performance by FD-Net (and vice versa). These results straightforwardly represent the success of FD-Net vs. the competing methods. In terms of image quality, FD-Net dominates over the competing methods, including the supervised baseline. While S-Net matches FD-Net in terms of SSIM over the predicted image quality, it lags behind when it comes to PSNR. As for the predicted field quality, FD-Net is second only to the supervised baseline.

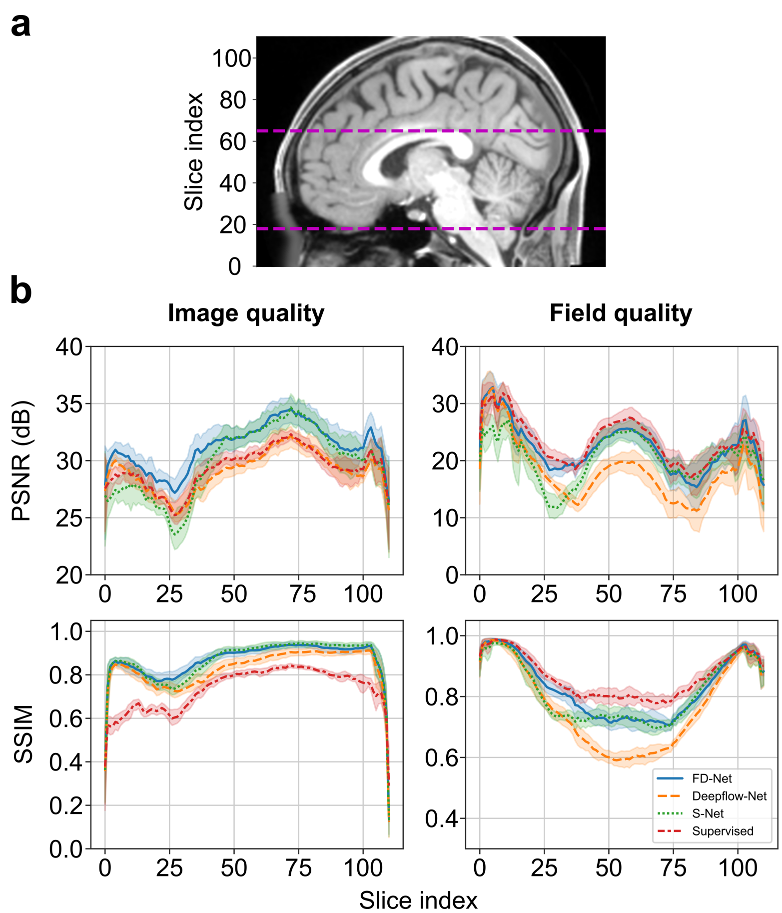

Figure 4.1: Slice-wise performance of FD-Net and competing methods. (a) A T1 weighted image registered to the b0 volume of one of the subjects, used to illustrate the anatomy corresponding to the slice index. Magenta dashed lines are used to indicate two slices of interest. (b) PSNR (top row) and SSIM (bottom row) results for the predicted image (left column) and field (right column), with respect to TOPUP. Color coded line plots of the 95% confidence interval around the mean for FD-Net and the competing methods are provided. Color codes are provided in the legend at the bottom right corner.

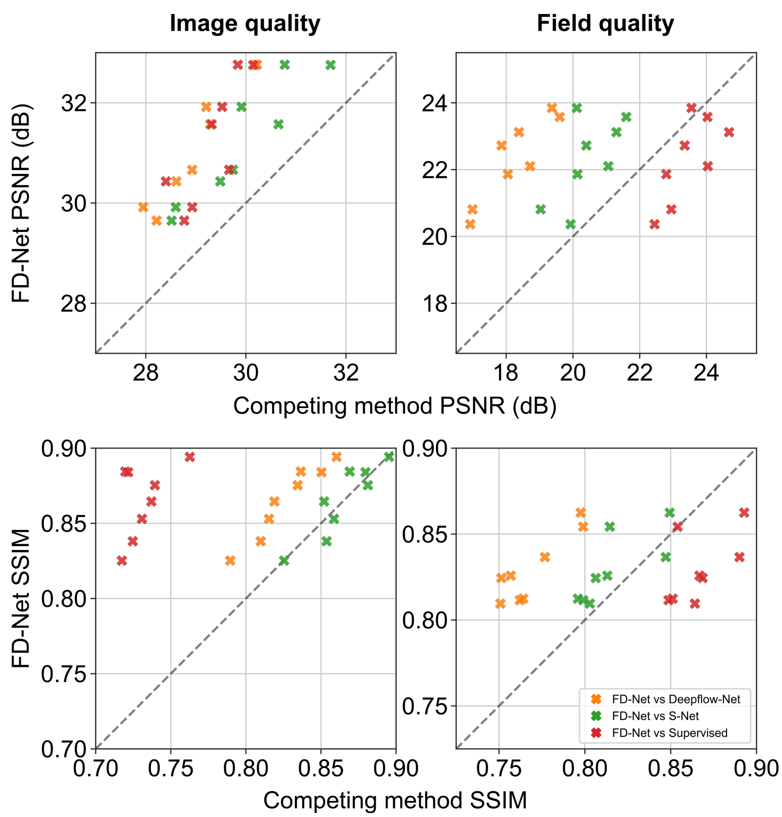

Figure 4.2: Subject-wise performance of FD-Net vs. competing methods. PSNR (top row) and SSIM (bottom row) results for the predicted image (left column) and field (right column), with respect to TOPUP, are provided as scatter plots for each of the 8 subjects reserved for testing. The y-axis represents FD-Net and the x-axis represents the competing method, color coded as per the legend at the bottom right corner. Each point denotes the overall mean performance of FD-Net vs. competing method for a given subject. Points above the dashed identity line represent the cases where FD-Net outperforms the respective competing method (and vice versa).

3. *Numerical Evaluation:* The quantitative results are arranged in Table 4.5 and Table 4.6 for PSNR and SSIM, respectively. These results summarize the performance of each method in terms of PSNR and SSIM computed over the predicted image and field across all the subjects. Image quality is boosted by 2.13dB PSNR/3.79% SSIM when compared to Deepflow-Net and 1.29dB PSNR/0.05% SSIM when compared S-Net. Field quality is boosted by 4.06dB PSNR/5.98% SSIM when compared to Deepflow-Net and 1.85dB PSNR/1.36% SSIM when compared to S-Net. Compared to the supervised baseline, FD-Net largely boosts performance in terms of image quality by 1.89dB PSNR/13.32% SSIM, with a small cost in field quality by 1.18dB PSNR/3.74% SSIM.

Table 4.5: PSNR performance of FD-Net and the competing methods. The leftmost column indicates the method considered, and the rightmost two columns indicate the PSNR performance of the predicted image and the field, respectively. PSNR is measured in dB with respect to the corresponding TOPUP results. The mean and standard deviation of PSNR across the subjects are reported. The best performing result in each column is highlighted in bold.

Method	Image quality	Field quality
Proposed FD-Net	**31.21 (2.76)**	22.30 (5.71)
Deepflow-Net	29.08 (2.33)	18.24 (6.48)
S-Net	29.92 (3.63)	20.45 (5.43)
Supervised baseline	29.32 (2.24)	**23.48 (5.06)**

Table 4.6: SSIM performance of FD-Net and the competing methods. The leftmost column indicates the method considered, and the rightmost two columns indicate the SSIM performance of the predicted image and the field, respectively. SSIM is provided as a percentage, measured with respect to the corresponding TOPUP results. The mean and standard deviation of SSIM across the subjects are reported. The best performing result in each column is highlighted in bold.

Method	Image quality	Field quality
Proposed FD-Net	**86.49 (11.53)**	82.96 (10.38)
Deepflow-Net	82.70 (11.72)	76.98 (14.19)
S-Net	86.44 (12.16)	81.60 (10.50)
Supervised baseline	73.17 (11.26)	**86.70 (7.63)**

4.2.2 Visual Slice Results

To compare the qualities of the predicted/corrected images and the predicted fields, visual results of the slices highlighted in magenta in Figure 4.1a are provided in Figure 4.3 for the lower brain slice and Figure 4.5 for the upper brain slice. These slices were chosen to represent the most and least challenging slices, corresponding to the dip and peak in PSNR in Figure 4.1b, respectively. The absolute error maps of the predicted image and the masked error maps of the predicted field are provided to facilitate comparison. Both error maps, as well as visual inspection of the predicted image and predicted field, point to FD-Net as the best performing method. Additionally, the forward-distorted images generated by FD-Net for the lower and upper brain slices are provided in Figure 4.4 and Figure 4.6, respectively. The display window on the error maps has been exaggerated to highlight the ability of the forward-distortion approach to constrain fidelity to the input EPI data.

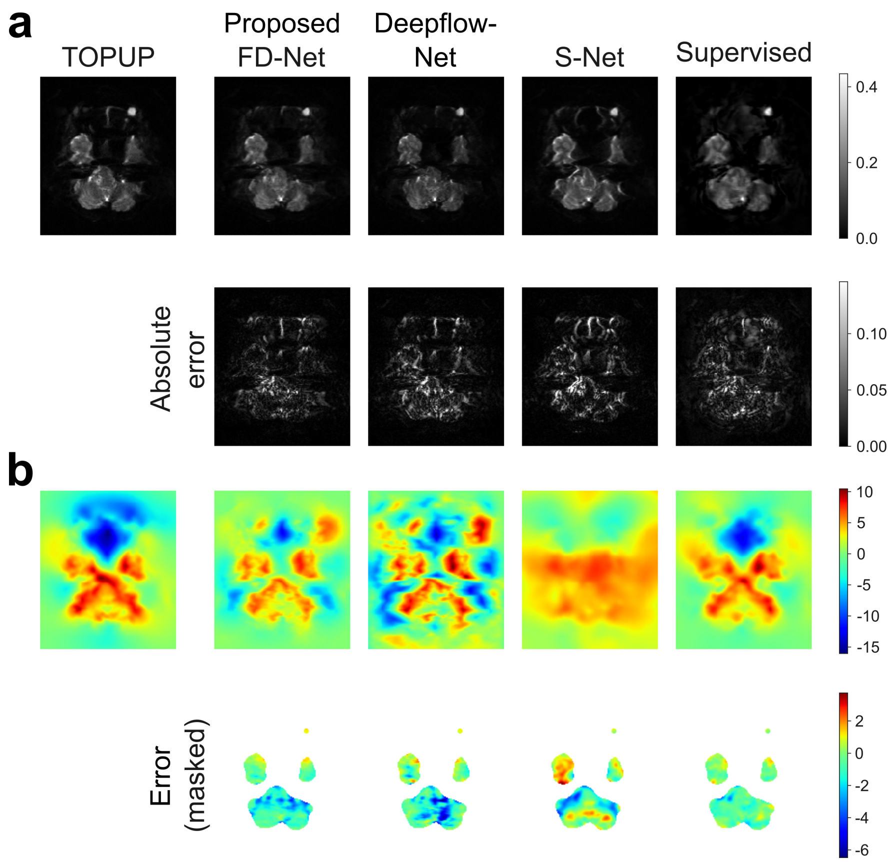

Figure 4.3: Visual results for FD-Net and competing methods from a lower brain slice. TOPUP results are provided in the leftmost column for reference, and are used to produce the respective error maps. (a) Predicted image results from each method are provided, with the absolute error maps with respect to TOPUP presented below. The error was scaled to a visibly discernible display window. (b) Predicted field results from each method are provided, with the masked error maps with respect to TOPUP presented below. Masking was performed via a median Otsu threshold over the TOPUP image to remove the background regions.

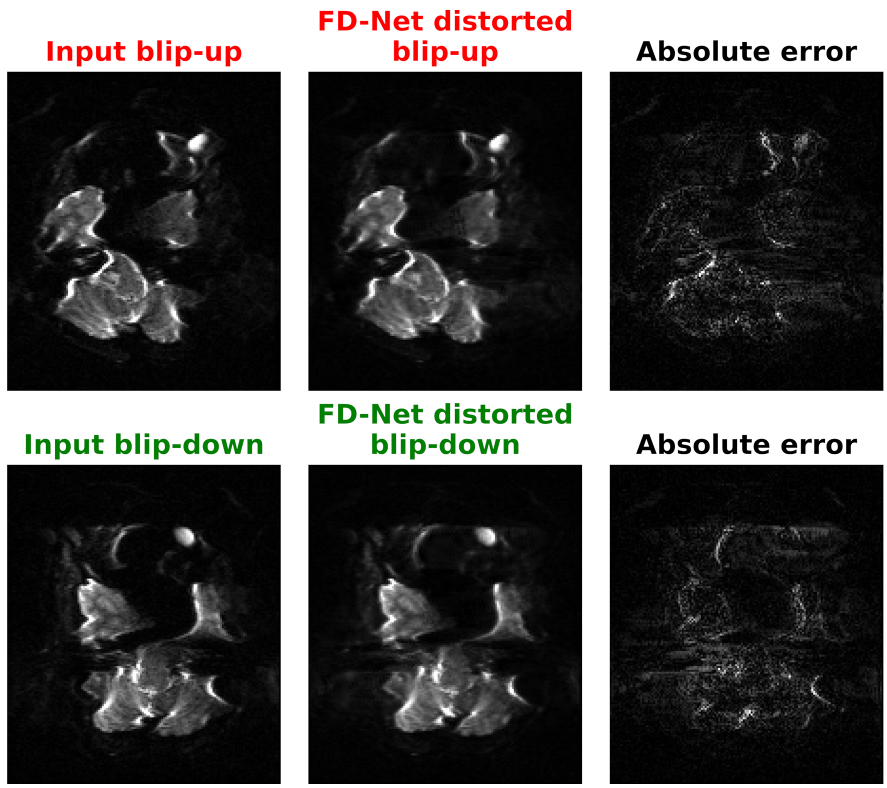

Figure 4.4: Visual results for the forward-distorted images in FD-Net from a lower brain slice. The input blip-up and blip-down EPI images are compared with the results of forward distortion in FD-Net, with the absolute error map provided in the far right column. The display window for the error is exaggerated in order to view the details of the error. The predicted image and predicted field for FD-Net for this slice are shown in Figure 4.3.

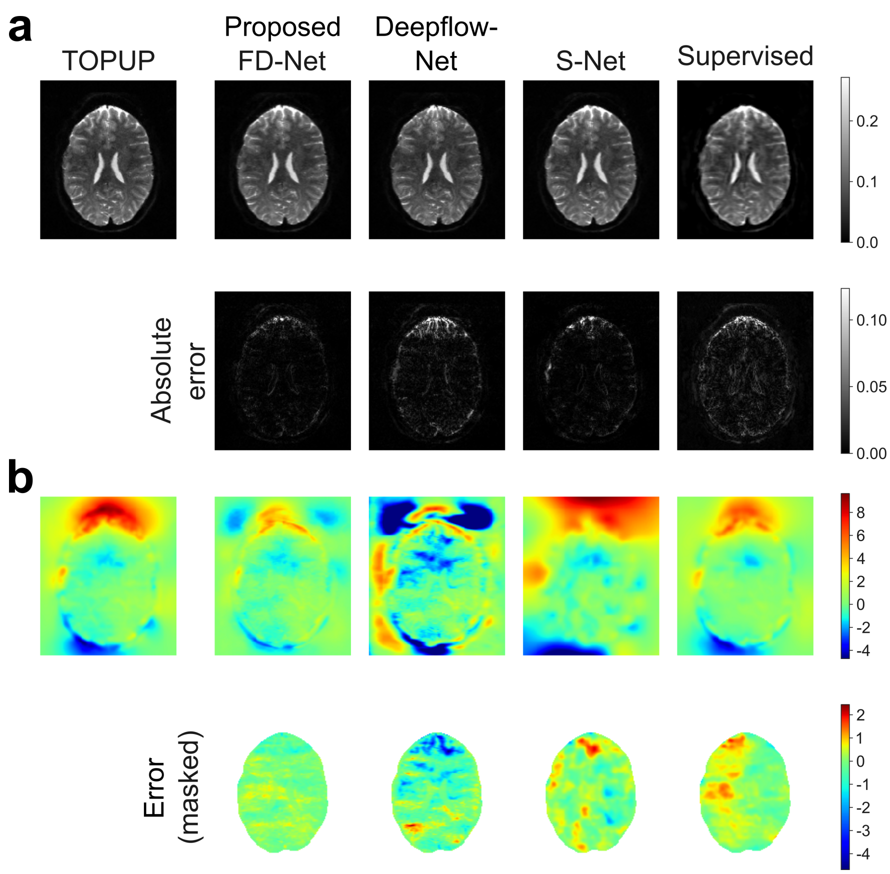

Figure 4.5: Visual results for FD-Net and competing methods from an upper brain slice. TOPUP results are provided in the leftmost column for reference, and are used to produce the respective error maps. (a) Predicted image results from each method are provided, with the absolute error maps with respect to TOPUP presented below. The error was scaled to a visibly discernible display window. (b) Predicted field results from each method are provided, with the masked error maps with respect to TOPUP presented below. Masking was performed via a median Otsu threshold over the TOPUP image to remove the background regions.

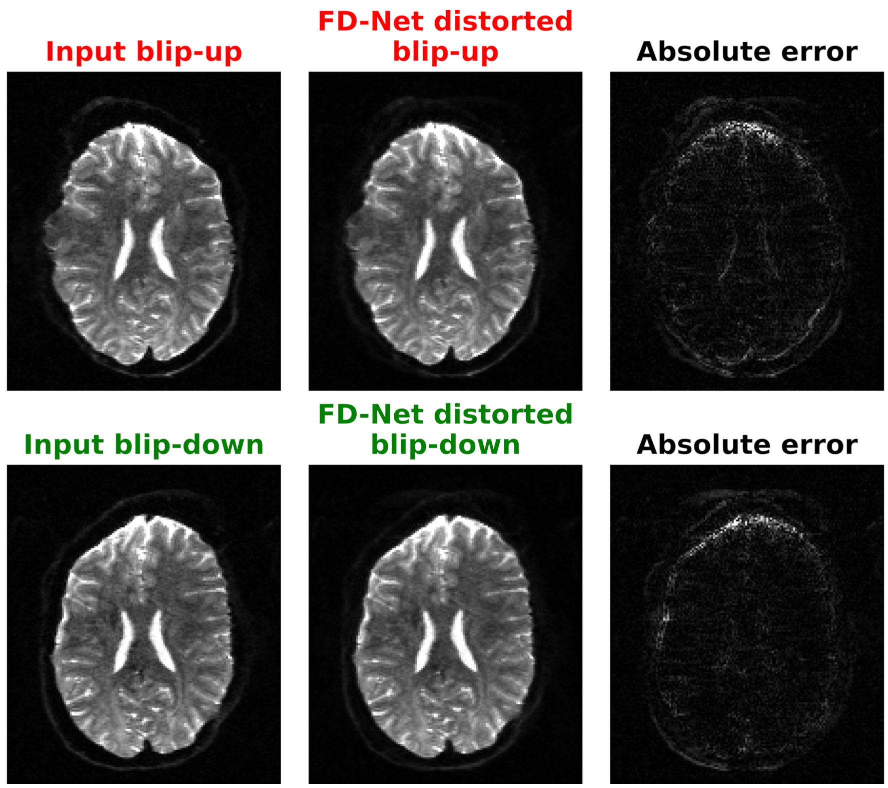

Figure 4.6: Visual results for the forward-distorted images in FD-Net from an upper brain slice. The input blip-up and blip-down EPI images are compared with the results of forward distortion in FD-Net, with the absolute error map provided in the far right column. The display window for the error is exaggerated in order to view the details of the error. The predicted image and predicted field for FD-Net for this slice are shown in Figure 4.5.

Chapter 5

Discussion and Conclusion

The results presented in Chapter 4 of this thesis demonstrate the success of the proposed FD-Net relative to the competing methods. Figure 4.1b shows that in terms of image quality (quantified via both PSNR and SSIM), FD-Net outperforms the other methods. This is especially true at the problematic lower slices (i.e., where large distortions are present). Regions of equitable performance tend to exhibit distortions that are not as severe, indicating a less challenging problem for the methods to solve.

FD-Net performs better than the other unsupervised methods in terms of field quality, as well. Even though the supervised baseline dominates in terms of the predicted field quality, FD-Net performs competitively despite being a fully unsupervised method. Since S-Net required a relatively large field smoothness regularization weight and is hence lacking in high frequency details especially at slices with less severe distortions, this might explain why its field similarity with respect to TOPUP is high in terms of SSIM in those slices.

Table 4.5 and Table 4.6 summarized the IQA results for PSNR and SSIM, respectively. FD-Net dominates in terms of image quality, and provides a noticeable improvement in field quality compared to Deepflow-Net and S-Net, while retaining largely similar field quality performance to that of the supervised baseline.

These numerical comparisons indicate that FD-Net is the best performing unsupervised method for correction. The boost in predicted image quality that FD-Net provides is greater than the decrease in predicted field quality with respect to the supervised baseline, demonstrating the advantages of the unsupervised nature of FD-Net.

Figure 4.2 compares the performance of FD-Net with the competing methods in a more direct fashion. Image quality comparisons show that FD-Net outperforms all competing methods, with S-Net only matching in terms of SSIM but lagging behind in PSNR, showing a discrepancy that can be attributed to intensity mismatches, possibly due to the lack of a density compensation scheme in S-Net. The field comparisons corroborate the superiority of FD-Net, only falling short to the supervised baseline. This latter result is to be expected, since the baseline was trained to directly fit the results from TOPUP. It should be noted that while this baseline is able to match the TOPUP field, it struggles to keep up in terms of predicted image quality.

The visual results in Figure 4.3 and Figure 4.5 show how FD-Net provides the best correction and field prediction performance in both a very challenging and a less challenging case. The predicted image has higher overall similarity to the TOPUP corrected image, with less artifacts present than the other methods. The field results also demonstrate how FD-Net produces the highest fidelity field, with smoothness and details preserved in a coherent manner. Figure 4.4 and Figure 4.6 also show the success of the forward-distortion approach in enforcing similarity to the input blip-up and blip-down EPI images.

The following is a more detailed commentary on the performance of each method with respect to FD-Net:

- Deepflow-Net suffers from deformities and mismatches in the correction, potentially due to its density compensation scheme depending on inverting the density pileup map. It is not straightforward how the cutoff should be decided when the density pileup map is in the range $(0, 1)$. This issue can cause incoherent modulation and therefore misguidance of the correction process. This problem can also explain why the generated field is perceptively less uniform, since the modulation could be manipulated by the network via the swing of the field amount. Overall, the effect potentially leads to a "tug of war" between the field generation and correction procedures.

- S-Net suffers from an underperformance issue, especially at the lower/more severely distorted slices. This issue also sheds light as to why the regularization parameter for field smoothness had to be set to a large amount. The lack of a density compensation scheme misguides the network, since the LCC loss is more tolerant of intensity mismatches. This tolerance gives the impression that the correction is performing well, while in reality it suffers from hyperintensities and inconsistencies in correcting the reversed PE images.

- The supervised baseline provides a closer match to the TOPUP field. However, visual comparison with the results of FD-Net reveals that this difference is minuscule given FD-Net's unsupervised nature. The boost in image quality over this baseline indicates that the unsupervised paradigm with the K-Unit approach in FD-Net is more suitable for performance generalization.

Due to the forward-distortion emphasis of FD-Net, similarity matching can be enforced directly in the same EPI space for each PE direction, as compared to the different corrected image spaces of the unwarping based correction approaches. This observation is especially true for multimodal image similarity measures, such as LCC of S-Net. While S-Net uses LCC for image matching, even though it is more robust to intensity variations [77], it leads to insufficient correction where large displacements occur.

The K-matrix formulation, as discussed earlier, also provides a straightforward way of emulating distortions without the issues of density compensation. Deepflow-net, while implementing a strategy for addressing pileups, relies on *estimating* the pileups by smoothed linear interpolation of the grid point density.

This process requires blurring the sampling density and multiplying by its inverse, introducing issues of scaling and parameter choice for the blurring.

Moreover, the global sharing of the parameters results in a natural regularization effect and, if desired or necessary, instance-specific optimization can be performed to adapt to new examples that might be very different from the previous training data [78].

Last but not least, FD-Net presents an additional potential advantage: since the forward-distortion "heads" and their outputs (Figure 3.1b-c) are not essential for delivering the predicted image and field, they can be omitted once training is completed and FD-Net is ready to be deployed for correction. This adaptation can simplify implementation of FD-Net in practice by limiting it to a purely convolutional neural network setup, making it more readily available for applications such as scanner-side correction with an optimized hardware setup.

In conclusion, the proposed FD-Net provides rapid correction of susceptibility artifacts in EPI, while maintaining high performance in terms of both image and field quality. This unsupervised deep learning approach provides results comparable to that of TOPUP, by predicting a self-consistent image and field. Our results indicate that the forward-distortion model methodizes a better-conditioned problem by enforcing consistency to measurement data, when compared to the unwarping based correction approaches. Therefore, FD-Net provides a novel paradigm for formulating the susceptibility artifact correction problem that better constrains fidelity to the measurement data.

Bibliography

[1] P. Mansfield, "Multi-planar image formation using NMR spin echoes," *Journal of Physics C: Solid State Physics*, vol. 10, no. 3, pp. L55–L58, 1977.

[2] S. J. Holdsworth and R. Bammer, "Magnetic Resonance Imaging Techniques: fMRI, DWI, and PWI," *Seminars in Neurology*, vol. 28, no. 4, pp. 395–406, 2008.

[3] R. Deichmann, J. A. Gottfried, C. Hutton, and R. Turner, "Optimized EPI for fMRI studies of the orbitofrontal cortex," *NeuroImage*, vol. 19, no. 2, pp. 430–441, 2003.

[4] K. M. Lüdeke, P. Röschmann, and R. Tischler, "Susceptibility artefacts in NMR imaging," *Magnetic Resonance Imaging*, vol. 3, no. 4, pp. 329–343, 1985.

[5] H. Chang and J. M. Fitzpatrick, "A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities," *IEEE Transactions on Medical Imaging*, vol. 11, no. 3, pp. 319–329, 1992.

[6] J.-D. Tournier, S. Mori, and A. Leemans, "Diffusion tensor imaging and beyond," *Magnetic Resonance in Medicine*, vol. 65, no. 6, pp. 1532–1556, 2011.

[7] P. Jezzard, "Correction of geometric distortion in fMRI data," *NeuroImage*, vol. 62, no. 2, pp. 648–651, 2012.

[8] D. Gallichan, "Diffusion MRI of the human brain at ultra-high field (UHF): A review," *NeuroImage*, vol. 168, pp. 172–180, 2018.

[9] J. L. R. Andersson, S. Skare, and J. Ashburner, “How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging,” *NeuroImage*, vol. 20, no. 2, pp. 870–888, 2003.

[10] D. Holland, J. M. Kuperman, and A. M. Dale, “Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging,” *NeuroImage*, vol. 50, no. 1, pp. 175–183, 2010.

[11] P. S. Morgan, R. W. Bowtell, D. J. O. McIntyre, and B. S. Worthington, “Correction of spatial distortion in EPI due to inhomogeneous static magnetic fields using the reversed gradient method,” *Journal of Magnetic Resonance Imaging*, vol. 19, no. 4, pp. 499–507, 2004.

[12] M. S. Graham, I. Drobnjak, M. Jenkinson, and H. Zhang, “Quantitative assessment of the susceptibility artefact and its interaction with motion in diffusion MRI,” *PLOS ONE*, vol. 12, no. 10, pp. 1–25, 2017.

[13] S. Ardekani and U. Sinha, “Geometric distortion correction of high-resolution 3 T diffusion tensor brain images,” *Magnetic Resonance in Medicine*, vol. 54, no. 5, pp. 1163–1171, 2005.

[14] K. V. Embleton, H. A. Haroon, D. M. Morris, M. A. L. Ralph, and G. J. M. Parker, “Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes,” *Human Brain Mapping*, vol. 31, no. 10, pp. 1570–1587, 2010.

[15] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. J. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, R. K. Niazy, J. Saunders, J. Vickers, Y. Zhang, N. De Stefano, J. M. Brady, and P. M. Matthews, “Advances in functional and structural MR image analysis and implementation as FSL,” *NeuroImage*, vol. 23, Supplement 1, pp. S208–S219, 2004.

[16] L. Ruthotto, S. Mohammadi, C. Heck, J. Modersitzki, and N. Weiskopf, “Hyperelastic Susceptibility Artifact Correction of DTI in SPM,” in *Bildverarbeitung für die Medizin 2013*, pp. 344–349, Springer Berlin Heidelberg, 2013.

[17] J. Modersitzki, *FAIR : flexible algorithms for image registration.* Society for Industrial & Applied Mathematics, 2009.

[18] F. Patzig, T. Mildner, T. Schlumm, R. Müller, and H. E. Möller, "Deconvolution-based distortion correction of EPI using analytic single-voxel point-spread functions," *Magnetic Resonance in Medicine*, vol. 85, no. 5, pp. 2445–2461, 2021.

[19] M. Jenkinson, J. L. Wilson, and P. Jezzard, "Perturbation method for magnetic field calculations of nonconductive objects," *Magnetic Resonance in Medicine*, vol. 52, no. 3, pp. 471–477, 2004.

[20] J. L. R. Andersson, "Chapter 4 - Geometric Distortions in Diffusion MRI," in *Diffusion MRI (Second Edition)*, pp. 63–85, Academic Press, 2014.

[21] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," 2015.

[22] S. T. M. Duong, S. L. Phung, A. Bouzerdoum, and M. M. Schira, "An unsupervised deep learning technique for susceptibility artifact correction in reversed phase-encoding EPI images," *Magnetic Resonance Imaging*, vol. 71, pp. 1–10, 2020.

[23] B. Zahneisen, K. Baeumler, G. Zaharchuk, D. Fleischmann, and M. Zeineh, "Deep flow-net for EPI distortion estimation," *NeuroImage*, vol. 217:116886, 2020.

[24] Y. Kim, S. Wiseman, A. C. Miller, D. Sontag, and A. M. Rush, "Semi-Amortized Variational Autoencoders," 2018.

[25] M. Poustchi-Amin, S. A. Mirowitz, J. J. Brown, R. C. McKinstry, and T. Li, "Principles and Applications of Echo-planar Imaging: A Review for the General Radiologist," *RadioGraphics*, vol. 21, no. 3, pp. 767–779, 2001.

[26] D. Le Bihan, J.-F. Mangin, C. Poupon, C. A. Clark, S. Pappata, N. Molko, and H. Chabriat, "Diffusion tensor imaging: Concepts and applications," *Journal of Magnetic Resonance Imaging*, vol. 13, no. 4, pp. 534–546, 2001.

[27] M. Drake-Pérez, J. Boto, A. Fitsiori, K. Lovblad, and M. I. Vargas, "Clinical applications of diffusion weighted imaging in neuroradiology," *Insights Imaging*, vol. 9, no. 4, pp. 535–547, 2018.

[28] H. J. Kim, C. G. Choi, D. H. Lee, J. H. Lee, S. J. Kim, and D. C. Suh, "High-b-Value Diffusion-Weighted MR Imaging of Hyperacute Ischemic Stroke at 1.5T," *American Journal of Neuroradiology*, vol. 26, no. 2, pp. 208–215, 2005.

[29] R. G. González, P. W. Schaefer, F. S. Buonanno, L. H. Schwamm, R. F. Budzik, G. Rordorf, B. Wang, A. G. Sorensen, and W. J. Koroshetz, "Diffusion-weighted MR imaging: diagnostic accuracy in patients imaged within 6 hours of stroke symptom onset," *Radiology*, vol. 210, no. 1, pp. 155–162, 1999.

[30] I. Caglic and T. Barrett, "Diffusion-weighted imaging (DWI) in lymph node staging for prostate cancer," *Translational Andrology and Urology*, vol. 7, no. 5, pp. 814–823, 2018.

[31] B. Taouli, V. Vilgrain, E. Dumont, J.-L. Daire, B. Fan, and Y. Menu, "Evaluation of Liver Diffusion Isotropy and Characterization of Focal Hepatic Lesions with Two Single-Shot Echo-planar MR Imaging Sequences: Prospective Study in 66 Patients," *Radiology*, vol. 226, no. 1, pp. 71–78, 2003.

[32] M. Cova, E. Squillaci, F. Stacul, G. Manenti, S. Gava, G. Simonetti, and R. Pozzi-Mucelli, "Diffusion-weighted MRI in the evaluation of renal lesions: preliminary results," *The British Journal of Radiology*, vol. 77, no. 922, pp. 851–857, 2004.

[33] E. Squillaci, G. Manenti, M. Cova, M. Di Roma, R. Miano, G. Palmieri, and G. Simonetti, "Correlation of Diffusion-weighted MR Imaging with Cellularity of Renal Tumours," *Anticancer Research*, vol. 24, no. 6, pp. 4175–4180, 2004.

[34] M. Eiber, K. Holzapfel, C. Ganter, K. Epple, S. Metz, H. Geinitz, H. Kübler, J. Gaa, E. J. Rummeny, and A. J. Beer, "Whole-body MRI including

diffusion-weighted imaging (DWI) for patients with recurring prostate cancer: Technical feasibility and assessment of lesion conspicuity in DWI," *Journal of Magnetic Resonance Imaging*, vol. 33, no. 5, pp. 1160–1170, 2011.

[35] K. Nasu, Y. Kuroki, S. Kuroki, K. Murakami, S. Nawano, and N. Moriyama, "Diffusion-weighted Single Shot Echo Planar Imaging of Colorectal Cancer Using a Sensitivity-encoding Technique," *Japanese Journal of Clinical Oncology*, vol. 34, no. 10, pp. 620–626, 2004.

[36] M. Afzali, T. Pieciak, S. Newman, E. Garyfallidis, E. Özarslan, H. Cheng, and D. K. Jones, "The sensitivity of diffusion MRI to microstructural properties and experimental factors," *Journal of Neuroscience Methods*, vol. 347, p. 108951, 2021.

[37] K. Abhinav, F.-C. Yeh, A. Mansouri, G. Zadeh, and J. C. Fernandez-Miranda, "High-definition fiber tractography for the evaluation of perilesional white matter tracts in high-grade glioma surgery," *Neuro-Oncology*, vol. 17, no. 9, pp. 1199–1209, 2015.

[38] K. Abhinav, F.-C. Yeh, S. Pathak, V. Suski, D. Lacomis, R. M. Friedlander, and J. C. Fernandez-Miranda, "Advanced diffusion MRI fiber tracking in neurosurgical and neurodegenerative disorders and neuroanatomical studies: A review," *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, vol. 1842, no. 11, pp. 2286–2297, 2014.

[39] W. Jost, *Diffusion in Solids, Liquids, Gases.* Academic Press, 1952.

[40] C. G. Koay and E. Özarslan, "Conceptual foundations of diffusion in magnetic resonance," *Concepts in Magnetic Resonance Part A*, vol. 42, no. 4, pp. 116–129, 2013.

[41] T. Schneider and C. A. M. Wheeler-Kingshott, "Chapter 3.2 - Q-Space Imaging: A Model-Free Approach," in *Quantitative MRI of the Spinal Cord*, pp. 146–155, Academic Press, 2014.

[42] H. Zhang, T. Schneider, C. A. Wheeler-Kingshott, and D. Alexander, "NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain," *NeuroImage*, vol. 61, no. 4, pp. 1000–1016, 2012.

[43] E. L. Hahn, "Spin Echoes," *Physical Review*, vol. 80, pp. 580–594, 1950.

[44] H. Y. Carr and E. M. Purcell, "Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments," *Physical Review*, vol. 94, no. 3, pp. 630–638, 1954.

[45] E. O. Stejskal and J. E. Tanner, "Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient," *The Journal of Chemical Physics*, vol. 42, no. 1, pp. 288–292, 1965.

[46] D. Le Bihan, R. Turner, C. T. W. Moonen, and J. Pekar, "Imaging of diffusion and microcirculation with gradient sensitization: Design, strategy, and significance," *Journal of Magnetic Resonance Imaging*, vol. 1, no. 1, pp. 7–28, 1991.

[47] S. T. Skare and R. Bammer, "EPI-Based Pulse Sequences for Diffusion Tensor MRI," in *Diffusion MRI: Theory, Methods, and Applications*, Oxford University Press, 2010.

[48] A. W. Anderson and J. C. Gore, "Analysis and correction of motion artifacts in diffusion weighted imaging," *Magnetic Resonance in Medicine*, vol. 32, no. 3, pp. 379–387, 1994.

[49] M. Descoteaux, *High Angular Resolution Diffusion Imaging (HARDI)*, pp. 1–25. John Wiley & Sons, Ltd, 2015.

[50] R. Turner and D. Le Bihan, "Single-shot diffusion imaging at 2.0 tesla," *Journal of Magnetic Resonance (1969)*, vol. 86, no. 3, pp. 445–452, 1990.

[51] G. C. McKinnon, "Ultrafast interleaved gradient-echo-planar imaging on a standard scanner," *Magnetic Resonance in Medicine*, vol. 30, no. 5, pp. 609–616, 1993.

[52] R. G. Nunes, P. Jezzard, T. E. J. Behrens, and S. Clare, "Self-navigated multishot echo-planar pulse sequence for high-resolution diffusion-weighted imaging," *Magnetic Resonance in Medicine*, vol. 53, no. 6, pp. 1474–1478, 2005.

[53] S. Moeller, S. Ramanna, C. Lenglet, P. K. Pisharady, E. J. Auerbach, L. Delabarre, X. Wu, M. Akcakaya, and K. Ugurbil, "Self-navigation for 3D multishot EPI with data-reference," *Magnetic Resonance in Medicine*, vol. 84, no. 4, pp. 1747–1762, 2020.

[54] F. Donato, D. N. Costa, Q. Yuan, N. M. Rofsky, R. E. Lenkinski, and I. Pedrosa, "Geometric Distortion in Diffusion-weighted MR Imaging of the Prostate—Contributing Factors and Strategies for Improvement," *Academic Radiology*, vol. 21, no. 6, pp. 817–823, 2014.

[55] S. C. Partridge, L. Singer, R. Sun, L. JWilmes, C. Klifa, C. D. Lehman, and N. M. Hylton, "Diffusion-weighted MRI: influence of intravoxel fat signal and breast density on breast tumor conspicuity and apparent diffusion coefficient measurements," *Magnetic Resonance Imaging*, vol. 29, no. 9, pp. 1215–1221, 2011.

[56] S. E. Williams, A. M. Heemskerk, E. B. Welch, K. Li, B. M. Damon, and J. H. Park, "Quantitative effects of inclusion of fat on muscle diffusion tensor MRI measurements," *Journal of Magnetic Resonance Imaging*, vol. 38, no. 5, pp. 1292–1297, 2013.

[57] J. J. Van Vaals and A. H. Bergman, "Optimization of eddy-current compensation," *Journal of Magnetic Resonance (1969)*, vol. 90, no. 1, pp. 52–70, 1990.

[58] J. Finsterbusch, "Eddy-current compensated diffusion weighting with a single refocusing RF pulse," *Magnetic Resonance in Medicine*, vol. 61, no. 3, pp. 748–754, 2009.

[59] T. G. Reese, O. Heid, R. M. Weisskoff, and V. J. Wedeen, "Reduction of eddy-current-induced distortion in diffusion MRI using a twice-refocused spin echo," *Magnetic Resonance in Medicine*, vol. 49, no. 1, pp. 177–182, 2003.

[60] M. H. Buonocore and L. Gao, "Ghost artifact reduction for echo planar imaging using image phase correction," *Magnetic Resonance in Medicine*, vol. 38, no. 1, pp. 89–100, 1997.

[61] N. kuei Chen and A. M. Wyrwicz, “Removal of EPI Nyquist ghost artifacts with two-dimensional phase correction,” *Magnetic Resonance in Medicine*, vol. 51, no. 6, pp. 1247–1253, 2004.

[62] M. Jaderberg, K. Simonyan, A. Zisserman, and K. Kavukcuoglu, “Spatial Transformer Networks,” in *Advances in Neural Information Processing Systems*, vol. 28, pp. 2017–2025, Curran Associates, Inc., 2015.

[63] A. Roche, G. Malandain, X. Pennec, and N. Ayache, “The correlation ratio as a new similarity measure for multimodal image registration,” in *Medical Image Computing and Computer-Assisted Intervention — MICCAI’98*, vol. 1496, pp. 1115–1124, Springer Berlin Heidelberg, 1998.

[64] M. Jenkinson and S. Smith, “A global optimisation method for robust affine registration of brain images,” *Medical Image Analysis*, vol. 5, no. 2, pp. 143–156, 2001.

[65] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, “Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images,” *NeuroImage*, vol. 17, no. 2, pp. 825–841, 2002.

[66] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, “SENSE: Sensitivity encoding for fast MRI,” *Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999.

[67] E. U. Saritas, C. H. Cunningham, J. H. Lee, E. T. Han, and D. G. Nishimura, “Dwi of the spinal cord with reduced fov single-shot epi,” *Magnetic Resonance in Medicine*, vol. 60, no. 2, pp. 468–473, 2008.

[68] B. A. Barlas, C. D. Bahadir, S. G. Kafali, U. Yilmaz, and E. U. Saritas, “Sheared two-dimensional radiofrequency excitation for off-resonance robustness and fat suppression in reduced field-of-view imaging,” *Magnetic Resonance in Medicine*, pp. 1–16, 2022.

[69] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil, “The WU-Minn Human Connectome Project: An overview,” *NeuroImage*, vol. 80, pp. 62–79, 2013.

[70] WU-Minn Consortium of the NIH Human Connectome Project, “1200 Subjects Reference Manual – Appendix I,” tech. rep., Human Connectome Project, 2017.

[71] F. Chollet *et al.*, “Keras,” 2015.

[72] M. Abadi *et al.*, “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems,” 2015. Software available from tensorflow.org.

[73] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” 2014.

[74] M. Staring, S. Klein, and J. P. W. Pluim, “A rigidity penalty term for nonrigid registration,” *Medical Physics*, vol. 34, no. 11, pp. 4098–4108, 2007.

[75] B. Fornberg, “Generation of finite difference formulas on arbitrarily spaced grids,” *Mathematics of Computation*, vol. 51, no. 184, pp. 699–706, 1988.

[76] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, “VoxelMorph: A Learning Framework for Deformable Medical Image Registration,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[77] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, “Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain,” *Medical Image Analysis*, vol. 12, no. 1, pp. 26–41, 2008.

[78] J. Marino, Y. Yue, and S. Mandt, “Iterative Amortized Inference,” 2018.