University of Wollongong

Research Online

An unsupervised deep learning technique for susceptibility artifact correction in reversed phase-encoding EPI images

Soan Duong
University of Wollongong, stmd795@uowmail.edu.au

Son Lam Phung
University of Wollongong, phung@uow.edu.au

Abdesselam Bouzerdoum
University of Wollongong, bouzer@uow.edu.au

Mark M. Schira
University of Wollongong, mschira@uow.edu.au

An unsupervised deep learning technique for susceptibility artifact correction in reversed phase-encoding EPI images

Abstract

© 2020 Elsevier Inc. Echo planar imaging (EPI) is a fast and non-invasive magnetic resonance imaging technique that supports data acquisition at high spatial and temporal resolutions. However, susceptibility artifacts, which cause the misalignment to the underlying structural image, are unavoidable distortions in EPI. Traditional susceptibility artifact correction (SAC) methods estimate the displacement field by optimizing an objective function that involves one or more pairs of reversed phase-encoding (PE) images. The estimated displacement field is then used to unwarp the distorted images and produce the corrected images. Since this conventional approach is time-consuming, we propose an end-to-end deep learning technique, named S-Net, to correct the susceptibility artifacts the reversed-PE image pair. The proposed S-Net consists of two components: (i) a convolutional neural network to map a reversed-PE image pair to the displacement field; and (ii) a spatial transform unit to unwarp the input images and produce the corrected images. The S-Net is trained using a set of reversed-PE image pairs and an unsupervised loss function, without ground-truth data. For a new image pair of reversed-PE images, the displacement field and corrected images are obtained simultaneously by evaluating the trained S-Net directly. Evaluations on three different datasets demonstrate that S-Net can correct the susceptibility artifacts in the reversed-PE images. Compared with two state-of-the-art SAC methods (TOPUP and TISAC), the proposed S-Net runs significantly faster: 20 times faster than TISAC and 369 times faster than TOPUP, while achieving a similar correction accuracy. Consequently, S-Net accelerates the medical image processing pipelines and makes the real-time correction for MRI scanners feasible. Our proposed technique also opens up a new direction in learning-based SAC.

Disciplines

Engineering | Science and Technology Studies

Publication Details

An Unsupervised Deep Learning Technique for Susceptibility Artifact Correction in Reversed Phase-encoding EPI Images

Soan T. M. Duong[a,*], Son L. Phung[a], Abdesselam Bouzerdoum[a,c], Mark M. Schira[b]

[a]*School of Electrical, Computer and Telecommunications Engineering, University of Wollongong*
[b]*School of Psychology, University of Wollongong, Australia*
[c]*ICT Division, College of Science and Engineering, Hamad Bin Khalifa University, Qatar*

Abstract

Echo planar imaging (EPI) is a fast and non-invasive magnetic resonance imaging technique that supports data acquisition at high spatial and temporal resolutions. However, susceptibility artifacts, which cause the misalignment to the underlying structural image, are unavoidable distortions in EPI. Traditional susceptibility artifact correction (SAC) methods estimate the displacement field by optimizing an objective function that involves one or more pairs of reversed phase-encoding (PE) images. The estimated displacement field is then used to unwarp the distorted images and produce the corrected images. Since this conventional approach is time-consuming, we propose an end-to-end deep learning technique, named S-Net, to correct the susceptibility artifacts the reversed-PE image pair. The proposed S-Net consists of two components: (i) a convolutional neural network to map a reversed-PE image pair to the displacement field; and (ii) a spatial transform unit to unwarp the input images and produce the corrected images. The S-Net is trained using a set of reversed-PE image pairs and an unsupervised loss function, without ground-truth data. For a new image pair of reversed-PE images, the displacement field and corrected images are obtained simultaneously by evaluating the trained S-Net directly. Evaluations on three different datasets demonstrate that S-Net can correct the susceptibility artifacts in the reversed-PE images. Compared with two state-of-the-art SAC methods (TOPUP and TISAC), the proposed S-Net runs significantly faster: 20 times faster than TISAC and 369 times faster than TOPUP, while achieving a similar correction accuracy. Consequently, S-Net accelerates the medical image processing pipelines and makes the real-time correction for MRI scanners feasible. Our proposed technique also opens up a new direction in learning-based SAC.

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

1. Introduction

Echo Planar Imaging (EPI) is the technique of choice for most functional magnetic resonance imaging (fMRI) and diffusion-weighted imaging (DWI) applications due to its fast imaging capability. Despite its popularity, EPI is sensitive to the local field inhomogeneities, which are caused by magnetic susceptibility differences of various imaged tissues, e.g. air versus bone, and fat versus blood [1, 2]. The field inhomogeneities affect the spatial encoding of the signal, resulting in intensity modulations and local image distortions (*i.e.* stretching and compressing) [3]. These distortions are known as susceptibility artifacts (SAs). They cause the misalignment to the underlying structural image, subsequently leading to incorrect localization of analysis results, such as the wrong activation patterns in the fMRI studies. SAs are more severe at high field strengths [4, 5], which have become widely used.

Several susceptibility artifact correction (SAC) methods rely on two reversed phase-encoding (PE) images, which are acquired using identical sequences but with opposite PE directions. The main idea is that the SAs appear inversely in the pair of reversed-PE images [6, 7, 8]; therefore, the middle version of the reversed-PE image pair is considered the corrected image [3]. Conventionally, the reversed-PE SAC methods involve two steps. In the first step, a displacement field along the PE direction is estimated by optimizing an objective function of one or more reversed-PE image pairs. Only the displacement field[1] along the PE direction is estimated since the displacements in the other directions are less significant [2, 9, 10]. In the second step, the estimated displacement field is used to unwarp the distorted images and produce the corrected images. This conventional approach is time-consuming, especially for input images with large sizes or severe displacements. Consequently, these SAC methods are unsuitable for time-sensitive applications, for example correction on an MRI scanner.

To reduce the processing time, we propose an unsupervised deep learning technique, called S-Net, for

---

*Corresponding author
    Email address:* stmd795@uowmail.edu.au (Soan T. M. Duong)

[1]For the rest of this paper, the term *displacement field* will refer to the displacement field in the PE direction.

correcting the susceptibility artifacts in 3D reversed-PE images. A convolutional neural network (CNN) is used to map a pair of reversed-PE images to the displacement field in the PE direction. Then, a differentiable spatial transform unit is used to unwarp the input (distorted) image pair via the predicted displacement field. The S-Net is trained in an end-to-end manner using a training set of reversed-PE image pairs. After training, correcting a new 3D reversed-PE image pair is achieved by simply evaluating the trained S-Net on the given input images. This approach, therefore, avoids the highly computational cost of the existing iterative optimization approaches.

The contributions of this paper are highlighted as follows.

1. We design a convolutional encoder-decoder network to map a 3D reversed-PE image pair to the displacement field. The network consists of an encoder for image downsampling and a decoder for image upsampling. The encoder uses a series of convolutional (conv) layers and leaky rectified linear units (LeakyReLUs) to extract hierarchical image contents. The decoder uses a series of conv layers, LeakyReLUs, and upsampling layers to recover the full-resolution image features and estimate the displacement field. A spatial transform unit is designed to unwarp the 3D input images along the PE direction and produce an output image pair. To the best of our knowledge, the proposed technique is the first attempt at using a learning-based approach with an convolutional encoder-decoder to correct the susceptibility artifacts.

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

3. We evaluate the performance of the proposed technique and compare it with existing SAC methods using three datasets. The datasets include one dataset acquired by our team using a 7 Tesla (T) scanner, and two public datasets acquired using a 3T scanner and published by the Human Connectome Project (HCP) [11]. The experimental results show that our unsupervised S-Net provides the corrected images which are comparable to results of state-of-the-art SAC methods while requiring fewer computational resources and no additional data, such as structural images.

The remainder of this paper is organized as follows. Section 2 presents the related work, and Section 3 introduces the proposed method. Section 4 presents

experiments and analysis of the proposed method and the related methods. Finally, Section 5 summarizes our work.

2. Related work

In this study, we investigate the reserved-PE based SAC because of its efficiency and popularity. This SAC approach is used to correct the fMRI and DWI data in the biggest MRI neuroimaging dataset - the HCP with 1200 subjects of multiple MRI modalities [12, 13]. Recall that the reserved-PE SAC first estimates the displacement field based on a pair of images acquired using an identical sequence but with opposite PE directions. The corrected images are then obtained by unwarping the distorted images via the estimated displacement field.

The conventional reversed-PE approach usually formulates SAC as an iterative optimization problem to estimate the displacement field. An overview of the conventional reversed-PE SAC methods is presented in Section 2.1. Besides, to the best of our knowledge, there have not had any reversed-PE SAC method using deep-learning based approach. Since the reversed-PE SAC can be cast as a symmetric non-rigid image registration problem, an overview of the learning-based non-rigid registration is presented in Section 2.2.

2.1. Conventional reversed-PE SAC

Several reversed-PE methods, as summarized in Table 1, have been proposed for correcting the susceptibility artifacts. The reversed-PE based SAC approach was initially introduced by Chang and Fitzpatrick for structural images [3]. They proposed a "cumulative line-integral" method to find the corresponding points, which are used to determine the displacement in two corresponding lines along the PE direction of the given reversed-PE images. An implementation of the method in [3] for 2D EPI was done by Bowtell et al. [17].

The corrections of the method proposed in [3] are not smooth since it estimates the displacement in each line along the PE direction independently, without considering the surrounding lines. To overcome this problem, Andersson et al. proposed an alternative method to estimate the displacement field [9]. They considered the displacement at a pixel as a function of discrete cosine basis functions to construct an objective function; this method is called TOPUP, and it is integrated into the FSL package[2].

Several methods have considered the reversed-PE based SAC as a registration problem. The two distorted reversed-PE images are registered so that their corrected versions are as similar as possible. Holland et al. proposed the first registration framework to correct SAs in the reversed-PE images [8]. To obtain the more realistic corrected images, Ruthotto et al. introduced

---

[2]Website: fsl.fmrib.ox.ac.uk/fsl/fslwiki/topup

Table 1: Representative reversed-PE based methods for correcting SAs.

Authors	Year	Description
Chang and Fitzpatrick	1992	Introduce the theoretical justification of the correction using reversed-PE images; correct each 1D image along the PE direction by finding pairs of corresponding points in the given two images.
Andersson et al.	2003	Model the displacement as a function of discrete cosine basis functions (called TOPUP).
Holland et al.	2010	Model the reversed-PE SAC as a diffusion registration problem.
Ruthotto et al.	2012	Introduce an additional non-linear regularizer into the diffusion regularized problem (called HySCO).
Irfanoglu et al.	2015	Use a T_{2w} image and the cross-correlation measure.
Duong et al.	2018	Use a T_{1w} image and the normalized gradient field measure.
Duong et al.	2020	Incorporate a T_{1w} image into the reversed-PE registration, and select the regularization parameters automatically by using the Bayesian optimization (called TISAC).

a regularization term, inspired by the hyper-elastic registration, into the registration framework [10]. This method is called HySCO, and its implementation is included in the SPM12 toolbox[3]. To improve the alignment of the corrected images to the anatomical image, Irfanoglu et al. added a regularization term based on the anatomical T_{2w} image into the reversed-PE SAC registration [14]. Instead of using the T_{2w} image, Duong et al. introduced the T_{1w} image into the regularization term, as the T_{1w} image is routinely acquired in a brain study [15]. They also improved their method to select automatically the regularization parameters by using the Bayesian optimization [16]; this method is called TISAC.

The main drawback of the reversed-PE based SAC methods is the high computational cost at test time because of the optimization step. For example, to correct a pair of reversed-PE images with a size of $192 \times 192 \times 36$ voxels, TOPUP takes approximately 12 minutes, and TISAC and HySCO take about a minute [16]. Furthermore, HySCO's accuracy depends on the regularization parameters, and TISAC requires an additional structural image.

2.2. Learning-based non-rigid image registration

Non-rigid image registration is typically formulated as an optimization problem to seek a non-linear correspondence for every pixel (or voxel) between a moving image and a fixed image. This approach could be slow due to the required iterative optimization. Learning-based registration is an approach to avoid the iterative optimization. The recent development in convolutional neural networks has shown remarkable successes in solving the image registration problem [18, 19, 20].

Several learning-based image registration methods have been proposed. Those methods usually consist of two parts: (i) a network to map the input image pair (fixed image and moving image) to displacement fields[4]; and (ii) a spatial transform unit to unwarp the input moving image via the predicted displacement fields. In term of ground-truth requirement for training, we group the registration methods into two categories: supervised registration and unsupervised registration.

The supervised registration methods usually train the mapping network by minimizing the difference between the output and desired displacement fields [21, 22, 23, 24, 25]. The desired displacement fields can be real [24]; however, most of them are synthesized using conventional registration methods, such as LDDMM shooting algorithm [22, 26] and SimpleITK[5] [23]. The mapping network can be designed as a patch-based CNN [21, 22, 23], a dual-stream CNN [24], or a convolutional encoder-decoder [25]. The supervised registration methods present a promising direction. However, artificial displacement fields are tedious to acquire and can restrict the types of deformations.

The unsupervised registration methods often combine the mapping network and the spatial transform unit into an end-to-end model, which takes an input image pair (fixed image and moving image) and produces a correction of the moving image. The network is trained with the loss function reflecting the smoothness of the displacement fields and the similarity between the input fixed image and the corrected moving image [27, 28, 29, 30]. The mapping network can be designed as a fully convolutional network [27, 28], or a convolutional encoder-decoder [29, 30]. The spatial transform unit can be implemented with deconvolutional operators [27], a cubic spline interpolation scheme [28], or a linear interpolation scheme [29, 30]. Note that the methods with the interpolation schemes are inspired by the spatial transformer network [31].

---

[3]Website: www.diffusiontools.com/documentation/hysco.html

[4]The image registration estimates the displacements in all dimensions for every pixel (voxel). In the 3D image registration, thus, three displacement fields are estimated.
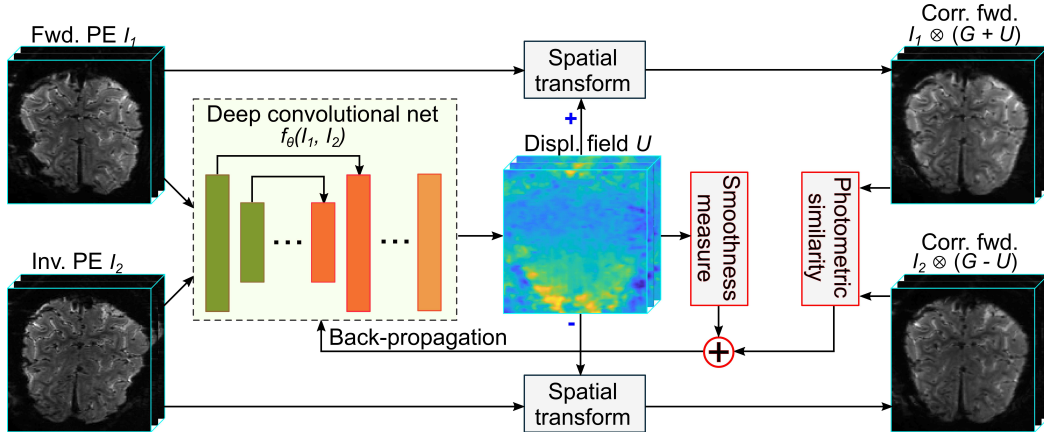
[5]http://www.simpleitk.org/

Figure 1: The proposed learning framework (S-Net) for correcting the SAs in reversed-PE images. The S-Net accepts a pair of reversed-PE images and produces the displacement field and the corrected images.

3. The proposed deep learning technique for SAC

This paper introduces a deep learning technique, called S-Net, for correcting a pair of 3D reversed-PE images. Fig. 1 illustrates an overview of the proposed S-Net. It consists of two parts: (i) a mapping function to estimate the displacement field U from a pair of 3D reversed-PE images I_1 and I_2; and (ii) a spatial transform unit to recover the corrected images by unwarping the input images with the estimated 3D displacement field. Note that the computations in the S-Net are performed in 3D.

The mapping function $f_{\boldsymbol{\theta}} : (I_1, I_2) \to U$ is modeled by a deep CNN, where $\boldsymbol{\theta}$ is the network parameters. Here, U is the 3D displacement field along the PE direction of image I_1 to the corrected image. Due to the inverse distortion property in the PE direction of the SAs, the displacement field of image I_2 to the corrected image is $-U$. The detail of the CNN architecture is presented in Section 3.1.

The spatial transform unit unwarps images I_1 and I_2 using the computed displacement field. The unwarping returns the corrected images, expressed as $\left[I_1 \otimes (G + U)\right]$ and $\left[I_2 \otimes (G - U)\right]$, where G is the identity transform (i.e. the regular grid), and $\otimes$ is the sampling operator. The detail of the spatial transform unit is presented in Section 3.2.

In the training phase, the network parameters $\boldsymbol{\theta}$ are learned to produce the best estimation of the displacement field and the corrected images. The displacement field is good if unwarping the input images using it produces the corrected images as similar to each other as possible. The goal of training is to minimize a loss function with respect to the network parameters $\boldsymbol{\theta}$ on a training dataset. In this work, the loss function is unsupervised. It captures the similarity of the corrected images and the smoothness of the displacement field. The detail of the unsupervised loss function is presented in Section 3.3.

Although the proposed S-Net is inspired by the VoxelMorph network presented in [29], we provide several new contributions. First, the loss function in [29] is

computed from the *input image* and the *output image* of VoxelMorph, while our loss function is computed from the *two output images* of S-Net. Second, VoxelMorph is designed for registering a moving T_1-weighted (T_{1w}) image to a fixed T_{1w} image, while S-Net is designed for correcting the susceptibility artifacts in a pair of EPI images acquired with reversed phase encoding. In VoxelMorph, the deformation is caused by affine transformations or anatomy differences between subjects. In contrast, in S-Net the deformation is caused by intrinsic magnetic susceptibility differences of tissues when putting in a magnetic fields.

3.1. CNN architecture for mapping

The mapping CNN architecture used in our method (see Fig. 2) is inspired by the U-Net [32], the VoxelMorph [29], and the DL-GP [33]. It accepts a two-channel image formed by concatenating two 3D reversed-PE images, and produces a 3D displacement field of the same size as the input images. The mapping network consists of an encoder (left side) and a decoder (right side). Both the encoder and decoder use a kernel size of 3 for their 3D conv layers.
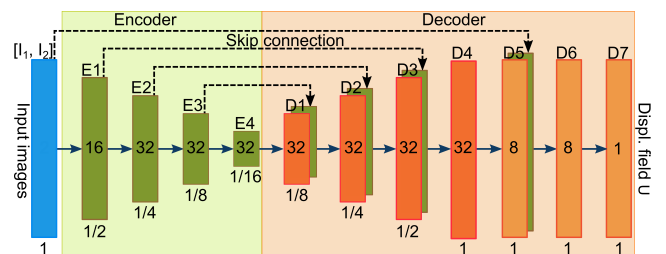

Figure 2: The convolutional neural network for mapping a pair of reversed-PE images to the displacement field. *Box*: output feature maps of a unit. *Number inside each box*: the number of feature maps in the unit. *Number below each box*: feature map size relative to the full input image size.

The encoder consists of four units E1, E2, E3 and E4; each unit consists of a conv layer and a LeakyReLU with a slope coefficient of 0.2. In the encoder, the conv layers

4

have a stride of 2 to reduce the size of their input feature maps by half. With this scheme, succeeding units of the encoder extract hierarchical features of the input image pair.

The decoder consists of seven units D1 to D7. Units D1, D2, D3, and D4 each consists of a conv layer, a LeakyReLU, and an upsampling layer. Each upsampling layer doubles the size of its feature maps, so the output feature maps of D4 have the same size as the input images I_1 and I_2. Units D5, D6, D7 each consists of a conv layer and a LeakyReLU. To retain more local information when upsampling and reduce the effects of vanishing gradients in training, skip connections are introduced in the network, as indicated in Fig. 2. For example, the 3D input images I_1 and I_2 are concatenated to the outputs of D5 before being fed to the next unit. The output feature maps of E1 are concatenated to the outputs of D3 before being fed to the next unit.

3.2. Spatial transformation unit for unwarping

This section presents the spatial transformation unit (STU), which is used to unwarp the distorted images and form the corrected images. The S-Net is trained by minimizing the loss function via a gradient-based algorithm, so the STU must be differentiable. Inspired by the spatial transformer network [31], we construct a differential STU unit, as illustrated in Fig. 3.
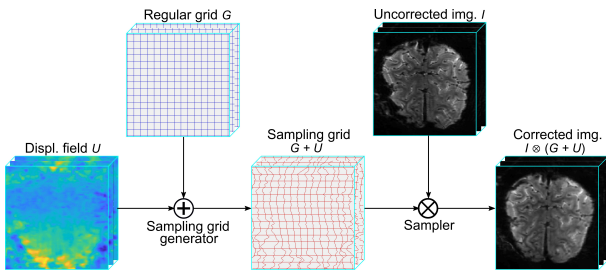

Figure 3: Spatial transformation unit for unwarping distorted image via the displacement field.

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

$$[I \otimes (G+U)](\mathbf{p}) = [1-(p'_x-q_{1,x})]I(\mathbf{q}_1)+[1-(q_{2,x}-p'_x)]I(\mathbf{q}_2), \quad (1)$$

where $\mathbf{q}_1$ and $\mathbf{q}_2$ are two neighbors of voxel $\mathbf{p}'$ along the first dimension with $q_{1,x} \leq p'_x < q_{2,x}$. The linear interpolation is selected due to its differentiability and computational efficiency.

3.3. Loss function for unsupervised network learning

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

$$\mathcal{L}(I_1, I_2, U) = \mathcal{L}_{\text{sim}}(E_1, E_2) + \lambda\, \mathcal{L}_{\text{smooth}}(U), \quad (2)$$

where λ is a positive regularization parameter.

In this study, we use the local cross-correlation (LCC) for $\mathcal{L}_{\text{sim}}$ for two reasons. First, LCC is suitable to measure the similarity of images with the same modality (i.e. the corrected images). Second, LCC is more robust to intensity variations found across scans and datasets [35].

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

$$\hat{X} = X - \bar{X}. \quad (3)$$

For a given voxel $\mathbf{p}$, let $W(\mathbf{p})$ denote the set of voxels in the $n \times n \times n$ volume centered on $\mathbf{p}$. For a pair of images E_1 and E_2, we compute an image C:

$$C(\mathbf{p}) = \frac{\left(\sum_{\mathbf{p}_i \in W(\mathbf{p})} \hat{E}_1(\mathbf{p}_i)\,\hat{E}_2(\mathbf{p}_i)\right)^2}{\sum_{\mathbf{p}_i \in W(\mathbf{p})} [\hat{E}_1(\mathbf{p}_i)]^2 \sum_{\mathbf{p}_i \in W(\mathbf{p})} [\hat{E}_2(\mathbf{p}_i)]^2}. \quad (4)$$

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

$$\text{LCC}(E_1, E_2) = \sum_{\mathbf{p}} C(\mathbf{p}), \quad (5)$$

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

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

$$\mathcal{L}_{\text{sim}}(E_1, E_2) = 1 - \text{LCC}(E_1, E_2). \quad (6)$$

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

$$\mathcal{L}_{\text{smooth}}(U) = \sum_{\mathbf{p} \in \Omega} \|\nabla U(\mathbf{p})\|^2, \quad (7)$$

5

where $\nabla U(\mathbf{p}) = \left(\frac{\partial_{U(\mathbf{p})}}{\partial_x}, \frac{\partial_{U(\mathbf{p})}}{\partial_y}, \frac{\partial_{U(\mathbf{p})}}{\partial_z} \right)$. The voxel difference is used to approximate $\nabla U(\mathbf{p})$, for example

$$\frac{\partial_{U(\mathbf{p})}}{\partial_x} \approx U(p_x + 1, p_y, p_z) - U(p_x, p_y, p_z). \qquad (8)$$

The proposed S-Net is trained by optimizing the loss function with respect to (w.r.t.) the trainable parameters $\boldsymbol{\theta}$. This is often done using stochastic optimization algorithms. In this paper, we adopt the *adaptive moment* estimation, known as Adam, [36]. This algorithm starts with an initial guess of $\boldsymbol{\theta}$. The next estimate of $\boldsymbol{\theta}$ is computed iteratively as

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \alpha \frac{m_t}{\sqrt{v_t} + \epsilon}, \qquad (9)$$

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

$$m_t = \frac{\beta_1 m_{t-1} + (1 - \beta_1) \nabla_{\boldsymbol{\theta}} \mathcal{L}_t}{1 - (\beta_1)^t}, \qquad (10)$$

$$v_t = \frac{\beta_2 v_{t-1} + (1 - \beta_2) [\nabla_{\boldsymbol{\theta}} \mathcal{L}_t]^2}{1 - (\beta_2)^t}, \qquad (11)$$

where hyper-parameters β_1 and β_2 are the exponential decay rates for the first and the second estimators. Kingma and Ba suggested the following hyper-parameter values: $\beta_1 = 0.9$ and $\beta_2 = 0.999$ [36].

4. Experiments and analysis

This section presents the experiments and analysis of the proposed S-Net. Section 4.1 describes the datasets used in the experiments, and Section 4.2 explains the experimental methods. Section 4.3 analyzes the regularization and learning rate parameters. Sections 4.4 and 4.5 compare the correction accuracy and processing time of the proposed method with other representative methods, respectively. Finally, Section 4.6 discusses the experimental results.

4.1. Description of datasets

The S-Net technique was evaluated using three datasets: fMRI-7T, fMRI-3T, and DWI-3T. The datasets are diverse in the acquisition sequence, modality, distortion property, field strength, resolution, image size, and dataset size. A summary of these datasets is presented in Table 2.

The fMRI-7T is a sub-millimeter dataset, which was acquired by our team for a retinotopic mapping fMRI study. It includes data of three healthy subjects acquired using a 7T scanner. Each subject was scanned in seven to ten *experimental runs* with the PE direction from left-to-right (LR). Each subject was also scanned in two *short runs* with the PE direction from right-to-left (RL), 20 seconds each. Each experimental run produced 183 or 187 images, and each short run produced 10 images. All images were motion-corrected using SPM12 tools [37]. To remove redundancy, we computed the mean image of each scan. The mean image of each experimental run was paired with the mean image of its respective inverse, resulting in 25 reversed-PE image pairs in total.

The fMRI-3T is a subset of the unpreprocessed 3T fMRI data in the public HCP dataset for studying functional connectivity of the human brain [12]. It includes data of 182 healthy subjects from the group *Subjects with 7T MR Session Data*. Each subject was scanned in 11 to 18 runs, using a customized Siemens 3T scanner. Each run contained two spin-echo images acquired by an identical sequence but with the PE direction alternated between RL and LR.

The DWI-3T is a subset of the unpreprocessed 3T DWI data in the public HCP dataset for reconstructing the complex axonal fiber architecture [12, 38]. It includes

Table 2: A summary of the datasets used in the experiments.

Datasets	No. subjs.	Gender dist.		Age dist.		Image size	Resolution	Acq. sequences	Field strength	PE directions
fMRI-7T	3	Males:	3	26-30 years:	1	$192 \times 192 \times 48$	$0.833 \times 0.833 \times 0.810$ mm^3	3D-GRE EPI (WIP1080 [34])	7T	LR and RL
				31-35 years:	1					
				over 36 years:	1					
fMRI-3T	182	Males:	72	22-25 years:	24	$90 \times 104 \times 72$	$2 \times 2 \times 2$ mm^3	Single-band 2D spin-echo EPI	3T	LR and RL
				26-30 years:	85					
		Females:	110	31-35 years:	71					
				over 36 years:	2					
DWI-3T	180	Males:	71	22-25 years:	23	$144 \times 168 \times 111$	$1.25 \times 1.25 \times 1.25$ mm^3	Single-band 2D spin-echo EPI	3T	LR and RL
				26-30 years:	84					
		Females:	109	31-35 years:	71					
				over 36 years:	2					

Abbreviations: LR = left-to-right; RL = right-to-left.
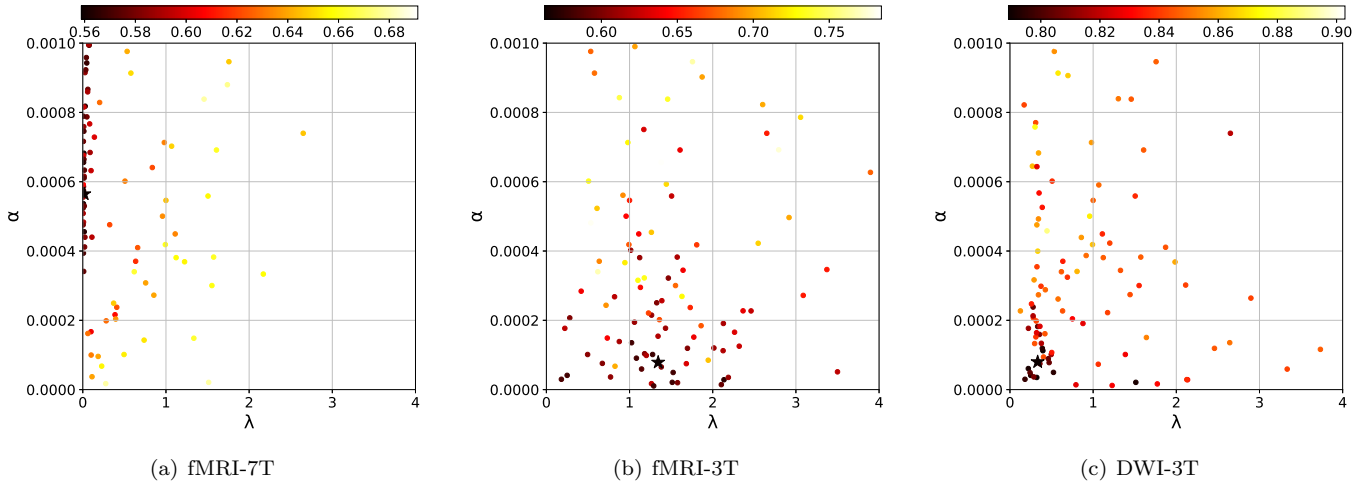
Figure 4: Validation loss of S-Nets versus the regularization parameter λ and learning rate α. The value pairs (α, λ) are found by the TPE algorithm. The color indicates the value of the validation loss; the star indicates the pair (α, λ) with the lowest validation loss.

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

Table 3: A summary of the training set and test set for each of the three datasets.

Datasets	Training set		Test set	
	No. subjects	No. pairs	No. subjects	No. pairs
fMRI-7T	2	13	1	9
fMRI-3T	156	2774	26	624
DWI-3T	150	436	30	90

4.2. Experimental methods

This section describes the S-Net implementation and the experimental setup. To implement the S-Net, we used Keras [39] with TensorFlow as the backend [40].

To evaluate the S-Net models, we split each dataset into a training set and a test set, as summarized in Table 3. The training set was used to select the hyper-parameters and train S-Net models, and the test set was used to evaluate the network performance.

To analyze the hyper-parameters, including the learning rate α of the Adam optimizer and the regularization parameter λ of the loss function in (2), we trained the S-Net model with multiple trials of hyper-parameter values. The trials were selected using the Tree of Parzen Estimator (TPE) algorithm from the Hyperopt library [41, 42, 43]. The objective function used in the TPE algorithm was the validation loss, which is the mean of the LCC measures between the pairs of output images from subsets of the training sets. For each dataset, the hyper-parameter values with the best validation loss were

selected for training the S-Net. Network training was done with a batch size of 3 for 1500 epochs.

To measure the correction accuracy of the proposed method, we computed the corrected images of the test sets using the S-Net and two state-of-the-art SAC methods, *i.e.* TOPUP [9] and TISAC [16]. The S-Net output images were then compared with the uncorrected images, and the output images produced by the two existing SAC methods. For accuracy comparison, we considered two types of similarity: (i) EPI similarity between the pair of reversed-PE images; and (ii) anatomical similarity between the EPI image and its corresponding T_{1w} image. The normalized mutual information (NMI) is used to measure the similarity. The NMI value range is from 0 (no mutual information) to 1 (complete correlation). In summary, we used two accuracy measures: *EPI-NMI* and *anatomical-NMI*.

To evaluate the processing speed, for each of the three SAC methods, we measured the times for two main tasks: (i) correcting the distorted image pair; and (ii) unwarping the distorted images given the displacement field. For the S-Net, image correction was done by evaluating the trained model with the input reversed-PE images, and unwarping was done via the spatial transformation unit. This experiment was conducted on the DWI-3T dataset. As shown in Table 2, this dataset has the largest image size and also the second largest number of subjects. All timings were reported on an Ubuntu 18.04.3 LTS workstation with an Intel Core i5-9600K Processor 3.6 GHz, 32 GB RAM, and an NVIDIA GeForce RTX 2080 GPU.

4.3. Analysis of hyper-parameters

This section analyzes the effects of two hyper-parameters of the model: the learning rate α of the Adam optimizer and the regularization parameter λ of the objective function in (2). Fig. 4 shows the validation loss for different values of α and λ for the three datasets. The
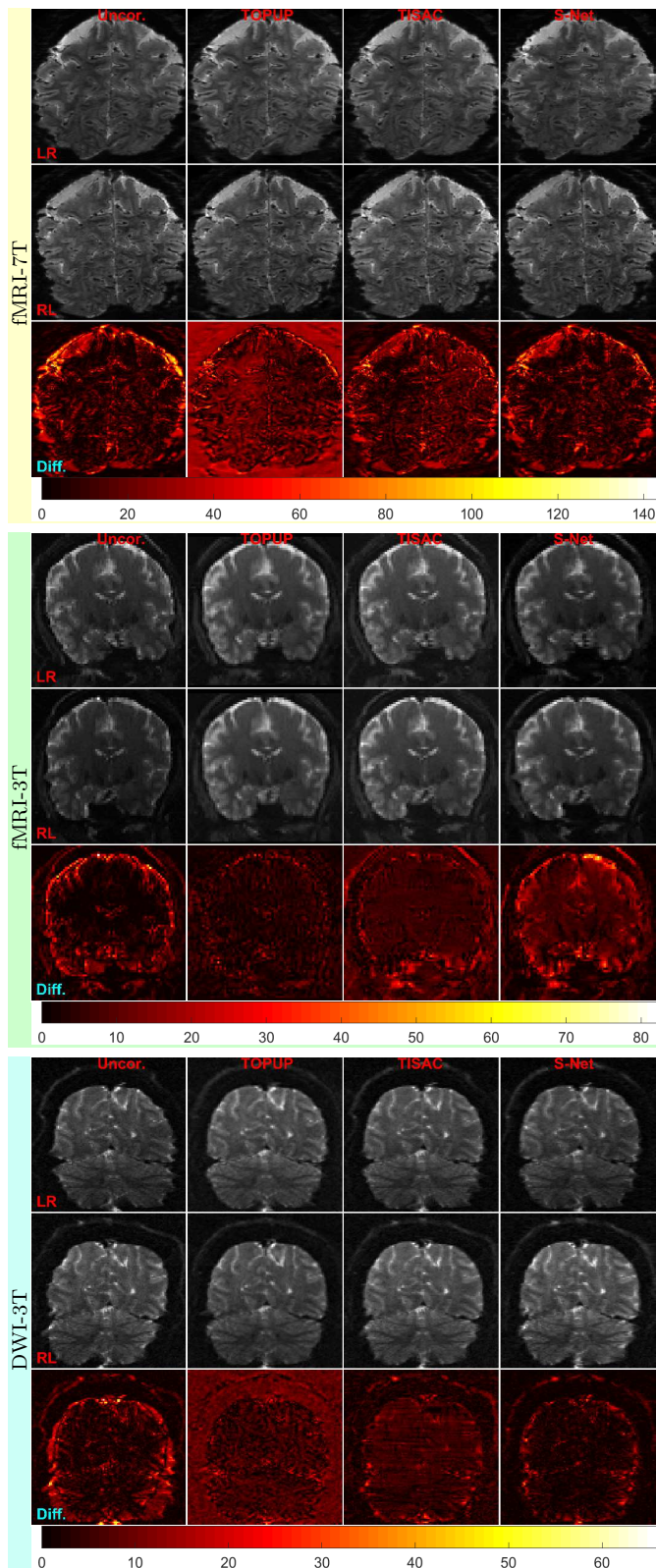
Figure 5: Representative visual results of SAC methods. *Column 1*: input uncorrected images. *Columns 2, 3, and 4*: output corrected images produced by TOPUP, TISAC, and S-Net, respectively. *Rows 1 and 2*: left-right and right-left phase-encoding EPI images, respectively. *Row 3*: absolute difference of LR and RL images.

dot represents the values of the pair (α, λ) produced by the TPE algorithm. The color of the dot encodes the validation loss: a dark color denotes a low validation loss, and a light color denotes a high validation loss. In each plot, the star ($*$) indicates the value pair (α, λ) that provides the lowest validation loss for the dataset.

Table 4: Hyper-parameter values used in training the S-Net models for the three datasets.

Hyper-params	fMRI-7T	fMRI-3T	DWI-3T
α	5.6e-4	7.87e-5	8e-5
λ	0.017	1.344	0.333

Fig. 4 shows that for the fMRI-7T dataset, the validation loss is small when $\alpha \in (0.0003, 0.001)$ and $\lambda \in (0, 0.0001)$. For the fMRI-3T dataset, the validation loss is small when $\alpha \in (0, 0.0001)$ and $\lambda \in (0, 2)$. For the DWI-3T dataset, the validation loss is small when $\alpha \in (0, 0.0001)$ and $\lambda \in (0, 1)$. These results show that there are several choices for (α, λ) that can give acceptable performances. The suitable values for (α, λ) as indicated by the TPE algorithm are summarized in Table 4.

4.4. Comparison with other methods in correction accuracy

This section presents both visual results and quantitative metrics of the proposed S-Net technique on the test sets. Fig. 5 shows one representative slice of uncorrected and corrected images from each of the three test sets. Each example includes the two reversed-PE images (Rows 1 and 2) and the absolute difference of the image pair (Row 3). It can be seen from Fig. 5 that S-Net removes the distortions in the uncorrected images. In addition, S-Net produces the comparable output images as TOPUP and TISAC.

Table 5 summarizes the statistical comparison of S-Net versus the uncorrected data (*i.e.* no correction) and the two SAC methods on the test sets. The p-values are computed by the paired t-tests on the NMI measures between the S-Net outputs and the compared images. The null hypothesis is $\mathcal{H}_0 : m_{\text{S-Net}} = m_{\text{other}}$. A p-value smaller than .05 indicates that the null hypothesis is rejected at a confidence level of 95%. In other words, there is a significant difference in the NMI measures between the S-Net outputs and the compared images.

Table 5 indicates that the S-Net technique improves both the EPI-NMI and anatomical-NMI measures of the uncorrected images significantly. The S-Net improves the NMI measures in most cases when compared to the other SAC methods. However, differences in the mean NMI measures between S-Net and the two SAC methods are small (see the visual comparisons in Fig. 6). In summary, S-Net is able to remove the geometric distortions in the reversed-PE images, and produce the corrected images at the same accuracy level of TOPUP and TISAC.

Table 5: Statistical comparisons of the NMI measures between the S-Net technique versus the uncorrected data and other two existing SAC methods on the test sets.

Datatypes	EPI-NMI						Anatomical-NMI					
	fMRI-7T		fMRI-3T		DWI-3T		fMRI-7T		fMRI-3T		DWI-3T	
	mean ± std	p-value	mean ± std	p-value	mean ± std	p-value	mean ± std	p-value	mean ± std	p-value	mean ± std	p-value
Uncorrected	0.302 ± 0.003	0.000	0.442 ± 0.023	0.000	0.282 ± 0.017	0.000	0.147 ± 0.014	0.000	0.733 ± 0.014	0.000	0.675 ± 0.012	0.000
TOPUP	0.943 ± 0.008	0.000	0.995 ± 0.001	0.000	0.984 ± 0.003	0.000	0.259 ± 0.002	0.000	0.917 ± 0.011	0.000	0.910 ± 0.013	0.000
TISAC	0.990 ± 0.001	0.000	0.992 ± 0.002	0.296	0.990 ± 0.001	0.000	0.250 ± 0.001	0.000	0.920 ± 0.013	0.000	0.915 ± 0.013	0.000
S-Net	0.976 ± 0.002		0.992 ± 0.000		0.991 ± 0.001		0.254 ± 0.003		0.924 ± 0.012		0.918 ± 0.013	

The null hypothesis $\mathcal{H}_0 : m_{\text{S-Net}} = m_{\text{other}}$. A p-value below .05 indicates that the null hypothesis is rejected at a confidence level of 95%.
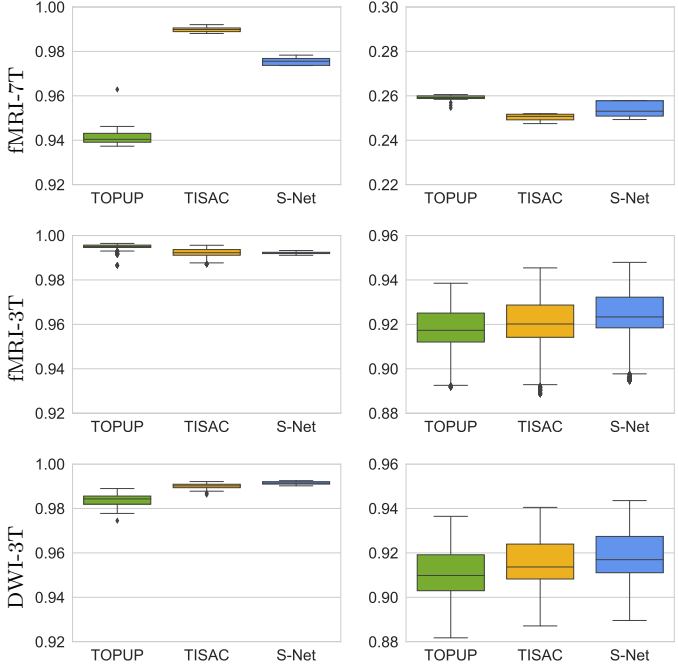
(a) In EPI-NMI measure (b) In anatomical-NMI measure

Figure 6: Visual comparisons of the accuracy between the S-Net technique versus other two existing SAC methods on the test sets. *Left column*: similarity between the corrected EPI images. *Right column*: similarity between the T_{1w} image and corrected EPI images. Because of differences in the datasets, the plots are drawn in different y-axis ranges for clarity.

4.5. Comparison with other methods in processing speed

This section compares the speed of the proposed S-Net method versus the two state-of-the-art SAC methods. Table 6 presents the average processing time of SAC methods for correcting a pair of reversed-PE images. To correct one reversed-PE image pair, S-Net required an average of 2.80 seconds, whereas TOPUP required an average of 1033.03 seconds and TISAC required an average of 56.49 seconds. Hence, S-Net is 369 times faster than TOPUP and 20 times than TISAC. Furthermore, when using the GPU, S-Net took only only 0.96 seconds to produce the corrected images. To the best of our knowledge, there is no GPU implementation for TOPUP and TISAC methods.

Table 7 shows the average runtimes taken to unwarp a pair of distorted images. To unwarp an image pair,

Table 6: Elapsed time (in second) of SAC methods for correcting a pair of reversed-PE images with a size of 144 × 168 × 111.

Processor	TOPUP (mean ± std.)	TISAC (mean ± std.)	S-Net (mean ± std.)
GPU	-	-	0.96 ± 0.07
CPU	1033.03 ± 382.03	56.49 ± 11.71	2.80 ± 0.10

Table 7: Elapsed time (in second) of SAC methods for unwarping separately two reversed-PE images with a size of 144 × 168 × 111.

Processor	TOPUP (mean ± std.)	TISAC (mean ± std.)	S-Net (mean ± std.)
GPU	-	-	1.33 ± 0.03
CPU	3.65 ± 0.06	3.40 ± 0.06	1.75 ± 0.09

TOPUP and TISAC required about 3.50 seconds, while S-Net on CPU required an average of 1.75 seconds, *i.e.* about two times faster than TOPUP and TISAC. S-Net on GPU required an average of 1.33 seconds, *i.e.* about 2.5 times faster than TOPUP and TISAC.

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

Table 8: Training time (in second) of the proposed S-Net model on different datasets.

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

For completeness, Table 8 summarizes the training time

of the proposed S-Net model for the three datasets. Note that this training time is incurred only once, and it does not affect the time for correcting images in the test phase. In contrast, as shown in Table 6 the existing methods require significant processing time (iterative optimization) for each new test image.

4.6. Discussion

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

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

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

5. Conclusion

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

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

Declarations of interest

The authors have declared that no competing interests exist.

References

[1] K. M. Ludeke, P. Roschmann, R. Tischler, Susceptibility artifacts in NMR imaging, Magn. Reson. Imaging 3 (1985) 329–343. doi:https://doi.org/10.1016/0730-725X(85)90397-2.

[2] F. Schmitt, Echo-Planar Imaging, volume 1, Academic Press, 2015, pp. 53–74.

[3] H. Chang, J. M. Fitzpatrick, A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities, IEEE Trans. Image Process. 11 (1992) 11. doi:https://doi.org/10.1109/42.158935.

[4] S. Ogawa, T. M. Lee, A. R. Kay, D. W. Tank, Brain magnetic resonance imaging with contrast dependent on blood oxygenation, in: Proc. Natl. Acad. Sci. U.S.A., 1990, pp. 9868–9872. doi:https://doi.org/10.1073/pnas.87.24.9868.

[5] J. R. Polimeni, V. Renvall, N. Zaretskaya, B. Fischl, Analysis strategies for high-resolution UHF-fMRI data, NeuroImage 168 (2018) 296–320. doi:https://doi.org/10.1016/j.neuroimage.2017.04.053.

[6] P. Jezzard, R. S. Balaban, Correction for geometric distortion in echo planar images from B0 field variations, Magn. Reson. Med. 34 (1995) 65–73. doi:https://doi.org/10.1002/mrm.1910340111.

[7] C. Hutton, A. Bork, O. Josephs, R. Deichmann, J. Ashburner, R. Turner, Image distortion correction in fMRI: a quantitative evaluation, NeuroImage 16 (2002) 217–240. doi:https://doi.org/10.1006/nimg.2001.1054.

[8] D. Holland, J. M. Kuperman, A. M. Dale, Efficient correction of inhomogeneous static magnetic field-induced distortion in echo planar imaging, NeuroImage 50 (2010) 175–184. doi:https://doi.org/10.1016/j.neuroimage.2009.11.044.

[9] J. L. R. Andersson, S. Skare, J. Ashburner, How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging, NeuroImage 20 (2003) 870–888. doi:https://doi.org/10.1016/S1053-8119(03)00336-7.

[10] L. Ruthotto, H. Kugel, J. Olesch, B. Fischer, J. Modersitzki, M. Burger, C. H. Wolters, Diffeomorphic susceptibility artifact correction of diffusion-weighted magnetic resonance images, Phys. Med. Biol. 57 (2012) 5715–5731. doi:https://doi.org/10.1088/0031-9155/57/18/5715.

[11] D. C. V. Essen, K. Ugurbil, E. Auerbach, D. Barch, T. E. J. Behrens, R. Bucholz, A. Chang, L. Chen, M. Corbetta, S. W. Curtiss, S. D. Penna, D. Feinberg, M. F. Glasser, N. Harel, A. C. Heath, L. Larson-Prior, D. Marcus, G. Michalareas, S. Moeller, R. Oostenveld, S. E. Petersen, F. Prior, B. L. Schlaggar, S. M. Smith, A. Z. Snyder, J. Xu, E. Yacoub, The human connectome project: a data acquisition perspective, NeuroImage 62 (2012) 2222–2231. doi:https://doi.org/10.1016/j.neuroimage.2012.02.018.

[12] D. C. V. Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, K. Ugurbil, The WU-Minn human connectome project: an overview, NeuroImage 80 (2013) 62–79. doi:https://doi.org/10.1016/j.neuroimage.2013.05.041.

[13] M. F. Glasser, S. N. Sotiropoulos, A. J. Wilson, T. S. Coalson, B. Fischl, J. L. Andersson, J. Xu, S. Jbabdi, M. Webster, J. R. Polimeni, D. C. V. Essen, M. Jenkinson, The minimal preprocessing pipelines for the human connectome project, NeuroImage 80 (2013) 105–124. doi:https://doi.org/10.1016/j.neuroimage.2013.04.127.

[14] M. O. Irfanoglu, P. Modia, A. Nayaka, E. B. Hutchinson, J. Sarllsc, C. Pierpaoli, DR-BUDDI (diffeomorphic registration for blip-up blip-down diffusion imaging) method for correcting echo planar imaging distortions, NeuroImage 106 (2015) 284–299. doi:https://doi.org/10.1016/j.neuroimage.2014.11.042.

[15] S. T. M. Duong, M. M. Schira, S. L. Phung, A. Bouzerdoum, H. G. B. Taylor, Anatomy-guided inverse-gradient susceptibility artefact correction method for high-resolution fmri, in: Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2018, pp. 786–790. doi:https://doi.org/10.1109/ICASSP.2018.8461352.

[16] S. T. M. Duong, S. L. Phung, A. Bouzerdoum, H. G. B. Taylor, A. M. Puckett, M. M. Schira, Susceptibility artifact correction for sub-millimeter fMRI using inverse phase encoding registration and T1 weighted regularization, J. Neurosci. Methods 336 (2020) 1–14. doi:https://doi.org/10.1016/j.jneumeth.2020.108625.

[17] R. Bowtell, D. McIntyre, M. Commandre, P. Glover, P. Mansfield, Correction of geometric distortion in echo planar images, in: Proc. Int. Soc. Magn. Reson., volume 2, 1994, p. 411.

[18] C. Doersch, A. Zisserman, Multi-task self-supervised visual learning, in: Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 2070–2079. doi:https://doi.org/10.1109/ICCV.2017.226.

[19] S. Khan, H. Rahmani, S. A. A. Shah, M. Bennamoun, G. Medioni, S. Dickinson, A Guide to Convolutional Neural Networks for Computer Vision, Synthesis Lectures on Computer Vision, Morgan & Claypool, 2018. doi:https://doi.org/10.2200/S00822ED1V01Y201712COV015.

[20] L. Jing, Y. Tian, Self-supervised visual feature learning with deep neural networks: a survey, arXiv preprint (2019) arXiv:1902.06162.

[21] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, D. Shen, Deformable image registration based on similarity-steered CNN regression, in: Proc. Med. Image. Comput. Assist. Interv., 2017, pp. 300–308. doi:https://doi.org/10.1007/978-3-319-66182-7_35.

[22] X. Yang, R. Kwitt, M. Styner, M. Niethammer, Quicksilver: fast predictive image registration a deep learning approach, NeuroImage 158 (2017) 378–396. doi:https://doi.org/10.1016/j.neuroimage.2017.07.008.

[23] H. Sokooti, B. D. d. Vos, F. Berendsen, B. P. F. Lelieveldt, I. Igum, M. Staring, Nonrigid image registration using multi-scale 3D convolutional neural networks, in: Proc. Med. Image. Comput. Assist. Interv., 2017, pp. 232–239. doi:https://doi.org/10.1007/978-3-319-66182-7_27.

[24] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, A. Kamen, Robust non-rigid registration through agent-based action learning, in: Proc. Med. Image. Comput. Assist. Interv., 2017, pp. 344–352. doi:https://doi.org/10.1007/978-3-319-66182-7_40.

[25] M.-M. Roh, M. Datar, T. Heimann, M. Sermesant, X. Pennec, SVF-Net: learning deformable image registration using shape matching, in: Proc. Med. Image. Comput. Assist. Interv., 2017, pp. 266–274. doi:https://doi.org/10.1007/978-3-319-66182-7_31.

[26] F.-X. Vialard, L. Risser, D. Rueckert, C. J. Cotter, Diffeomorphic 3D image registration via geodesic shooting using an efficient adjoint calculation, Int. J. Comput. Vis. 97 (2012) 229–241. doi:https://doi.org/10.1007/s11263-011-0481-8.

[27] H. Li, Y. Fan, Non-rigid image registration using self-supervised fully convolutional networks without training data, in: Proc. IEEE Int. Symp. Biomed. Imaging, 2018, pp. 1075–1078. doi:https://org.doi/10.1109/ISBI.2018.8363757.

[28] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, I. Igum, End-to-end unsupervised deformable image registration with a convolutional neural network, in: Deep Learn. Med. Image Anal. Multimodal Learn. Clin. Decis. Support, 2017, pp. 204–212. doi:https://doi.org/10.1007/978-3-319-67558-9_24.

[29] G. Balakrishnan, A. Zhao, M. R. Sabuncu, A. V. Dalca, J. Guttag, An unsupervised learning model for deformable medical image registration, in: Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., 2018, pp. 9252–9260. doi:https://doi.org/10.1109/CVPR.2018.00964.

[30] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, A. V. Dalca, VoxelMorph: a learning framework for deformable medical image registration, IEEE Trans. Med. Imaging 38 (2019) 1788–1800. doi:https://doi.org/10.1109/TMI.2019.2897538.

[31] M. Jaderberg, K. Simonyan, A. Zisserman, K. Kavukcuoglu, Spatial transformer networks, in: Proc. Int. Conf. Neural Inf. Process. Sys., volume 2, 2015, pp. 2017–2025.

[32] O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in: Proc. Med. Image. Comput. Assist. Interv., 2015, pp. 234–241.

[33] T. N. A. Nguyen, S. L. Phung, A. Bouzerdoum, Hybrid deep learning-gaussian process network for pedestrian lane detection in unstructured scenes, IEEE Trans. Neural Netw. Learn. Sys. (2020) 1–15. doi:https://doi.org/10.1109/TNNLS.2020.2966246.

[34] B. A. Poser, P. J. Koopmans, T. Witzel, L. L. Wald, M. Barth, Three dimensional echo-planar imaging at 7 Tesla, NeuroImage 51 (2010) 261–266. doi:https://doi.org/10.1016/j.neuroimage.2010.01.108.

[35] B. B. Avants, C. L. Epstein, M. Grossman, J. C. Gee, Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain, Med. Image. Anal. 12 (2008) 26–41. doi:https://doi.org/10.1016/j.media.2007.06.004.

[36] D. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint (2014) arXiv:1412.6980.

[37] W. Penny, K. Friston, J. Ashburner, S. Kiebel, T. Nichols, Statistical Parametric Mapping: The Analysis of Functional Brain, 1 ed., Academic Press, 2006.

[38] K. Uurbil, J. Xu, E. J. Auerbach, S. Moeller, A. T. Vu, J. M. Duarte-Carvajalino, C. Lenglet, X. Wu, S. Schmitter, P. F. V. d. Moortele, J. Strupp, G. Sapiro, F. D. Martino, D. Wang, N. Harel, M. Garwood, L. Chen, D. A. Feinberg, S. M. Smith, K. L. Miller, S. N. Sotiropoulos, S. Jbabdi, J. L. R. Andersson, T. E. J. Behrens, M. F. Glasser, D. C. V. Essen, E. Yacoub, Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome

Project, NeuroImage 80 (2013) 80–104. doi:https://doi.org/10.1016/j.neuroimage.2013.05.012.

[39] F. Chollet, et al., Keras, https://keras.io, 2015.

[40] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: large-scale machine learning on heterogeneous systems, https://www.tensorflow.org/, 2015.

Software available from tensorflow.org.

[41] J. Bergstra, R. Bardenet, Y. Bengio, B. Kgl, Algorithms for hyper-parameter optimization, in: Proc. Int. Conf. Neural Inf. Process. Sys., 2011, pp. 2546–2554.

[42] J. Bergstra, D. Yamins, D. D. Cox, Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures, in: Proc. Int. Conf. Machine Learning, 2013, pp. I–115–I–123.

[43] J. Bergstra, B. Komer, C. Eliasmith, D. Yamins, D. D. Cox, Hyperopt: a Python library for model selection and hyperparameter optimization, Comput. Sci. Discov. 8 (2015) 014008. doi:https://doi.org/10.1088/1749-4699/8/1/014008.