Article

Correcting Susceptibility Artifacts of MRI Sensors in Brain Scanning: A 3D Anatomy-guided Deep Learning Approach

Soan T. M. Duong [1,4,5]*, Son L. Phung [1], Abdesselam Bouzerdoum[1,3], Sui P. Ang[1], and Mark M. Schira [2]

1 School of Electrical, Computer and Telecommunications Engineering, University of Wollongong, Australia
2 School of Psychology, University of Wollongong, Australia
3 ICT Division, College of Science and Engineering, Hamad Bin Khalifa University, Qatar
4 Applied Science Division, VinBrain, VinGroup, Vietnam
5 Faculty of Information Technology, Le Quy Don Technical University, Vietnam
* Correspondence: stmd795@uowmail.edu.au

Abstract: Echo planar imaging (EPI), a fast magnetic resonance imaging technique, is a powerful tool in functional neuroimaging studies. However, susceptibility artifacts, which cause misinterpretations of brain functions, are unavoidable distortions in EPI. This paper proposes an end-to-end deep learning framework, named TS-Net, for susceptibility artifact correction (SAC) in a pair of 3D EPI images with reversed phase-encoding directions. The proposed TS-Net comprises a deep convolutional network to predict a displacement field in three dimensions to overcome the limitation of existing methods, which only estimate the displacement field along the dominant-distortion direction. In the training phase, anatomical T1-weighted images are leveraged to regularize the correction, but they are not required during the inference phase to make TS-Net more flexible for general use. The experimental results show that TS-Net achieves favorable accuracy and speed trade-off when compared with the state-of-the-art SAC methods, i.e. TOPUP, TISAC, and S-Net. The fast inference speed (less than a second) of TS-Net makes real-time SAC during EPI image acquisition feasible, and accelerates the medical image-processing pipelines.

Keywords: Susceptibility artifacts; deep learning; high-speed; echo planar imaging; reversed phase-encoding.

Citation: Duong, S. T. M; Phung, S. L.; Bouzerdoum, A.; Ang S. P.; Schira M. M, Correcting Susceptibility Artifacts of MRI Sensors in Brain Scanning: A 3D Anatomy-guided Deep Learning Approach. *Sensors* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Echo planar imaging is a fast magnetic resonance imaging (MRI) technique that has served as an important non-invasive tool in cognitive neuroscience [1]. EPI is widely used to record the functional magnetic resonance imaging (fMRI) data for studying human brain functions [2]. It is also the technique of choice to acquire the diffusion-weighted imaging (DWI) data for analyzing brain connection patterns [3]. Despite its popularity, EPI is prone to susceptibility artifacts (SAs) [4,5] and Eddy-current artifacts [6, 7], which consist of geometric distortions. The geometric distortions cause misalignments between the functional image and the underlying structural image, subsequently leading to errors in brain analysis, e.g. incorrect localization of neural activities in the functional brain studies. Therefore, an accurate geometric distortion correction method is crucial for applications that rely on EPI images.

In this study, we investigate the susceptibility artifact correction (SAC) as SAs are inevitable in EPI [5]. Interestingly, two EPI images, which are acquired using identical sequences but with reversed phase-encoding (PE) directions, have opposite patterns of geometric distortions caused by SAs[8,9]. Consequently, the middle version of the reversed-PE image pair is considered the distortion-free image. Chang and Fitzpatrick proposed to correct the SAs in two reversed-PE images by finding the corresponding points between two reversed-PE images; the corrected image was then formed by the mean intensity of the corresponding points [4]. Since displacements are estimated in lines along the PE direction independently, the estimated displacement field is not smooth, subsequently leading to unrealistic corrections. Andersson et al. proposed a method, called TOPUP, by modeling the displacement at each voxel as a function of discrete cosine basis functions [10]. This method estimates the entire *displacement field* along the PE direction, thereby avoiding the unsmooth problem.

Several reversed-PE based SAC methods have adopted an image registration approach, in which the corrected image is treated as the intermediate version of the two distorted input images. The two distorted reversed-PE images are transformed to the corrected image by an equal displacement amount but with the opposite directions. This registration approach for reversed-PE SAC was firstly proposed in [9]. Ruthotto et al. introduced a regularization term, inspired by the hyper-elastic registration, to constrain the displacement field in the registration framework, thereby achieving more realistic corrected images [11]. Hedouin et al. introduced the block-matching algorithm that estimates the displacement field at the block level of the given EPI image pair [12]. In another approach, Irfanoglu et al. introduced an anatomical regularizer based on the T2-weighted (T_{2w}) image to the registration framework so as to align better the corrected images to the underlying anatomical structure [13]. Duong et al. utilized T1-weighted (T_{1w}) for correction regularization as the T_{1w} images are routinely acquired in brain studies [14,15]; this method is called TISAC.

The above SAC methods require an iterative-optimization algorithm to estimate the displacement field and then compute the corrected images. This computation-intensive optimization step can take from one to 12 min, for an image pair of size $192 \times 192 \times 36$ voxels [15]. Recently, Duong et al. proposed an end-to-end deep learning framework, called S-Net, to map a pair of 3D input reversed-PE images to a displacement field in the phase-encoding direction, and provide the corrected image pair [16]. S-Net is trained using a set of reversed-PE image pairs. A new image pair is corrected by feeding the distorted image pair to the trained S-Net model directly, thereby reducing the processing time. The results of S-Net demonstrate the feasibility of using a deep network for the SAC problem. While providing a competitive correction accuracy, S-Net could still be improved in terms of correction accuracy, robustness to input image sizes, and imaging modalities.

To reduce computation time and increase robustness, existing SAC methods estimate the displacement field only along the phase-encoding direction (i.e. 1D distortion model). This is based on the fact that the distortions in the PE direction are prominent, whereas the distortions in the other directions are insignificant. In this study, we propose a generalized approach to enhance the correction accuracy by considering the distortions in all three directions (i.e. 3D distortion model). The 3D displacement field is predicted through a 3D convolutional encoder-decoder given a 3D reversed phase-encoding image pair. The convolutional network is trained end-to-end using the T_{1w} modality as an auxiliary condition. The proposed method is called anatomy-guided deep learning SAC, or TS-Net in which the letter "T" arises from T_{1w}.

The new contributions of this paper are highlighted as follows:

1. We design a deep convolutional network to estimate the 3D displacement field. The deep network is designed to make TS-Net robust to different sizes, resolutions, and modalities of the input image by using batch normalization (BN) layers and size-normalized layers.
2. We estimate the displacement field in *all* three dimensions instead of only along the phase-encoding direction. In other words, TS-Net predicts the displacement field that captures the 3D displacements for every voxel. This, to our knowledge, is a significant improvement compared to most existing SAC methods [10,16], which estimate the distortions only along the PE direction and ignore the distortions along with the other two directions.

Table 1: A summary of the datasets used in the experiments.

Datasets	No. subjs.	Gender distribution	Age distribution	Image size (voxels)	Resolution (mm^3)	Acquisition sequences	BW Hz/P_x	Field strength	PE directions
fMRI-3T	182	Males: 72 Females: 110	22-25 years: 24 26-30 years: 85 31-35 years: 71 over 36 years: 2	90 × 104 × 72	2 × 2 × 2	Multi-band 2D gradient-echo EPI, factor of 8	2290	3T	LR and RL
DWI-3T	180	Males: 71 Females: 109	22-25 years: 23 26-30 years: 84 31-35 years: 71 over 36 years: 2	144 × 168 × 111	1.25 × 1.25 × 1.25	Multi-band 2D spin-echo EPI, factor of 3	1488	3T	LR and RL
fMRI-7T	184	Males: 72 Females: 112	22-25 years: 24 26-30 years: 85 31-35 years: 73 over 36 years: 2	130 × 130 × 85	1.6 × 1.6 × 1.6	Multi-band 2D gradient-echo EPI, factor of 5	1924	7T	AP and PA
DWI-7T	178	Males: 69 Females: 109	22-25 years: 21 26-30 years: 85 31-35 years: 70 over 36 years: 2	200 × 200 × 132	1.05 × 1.05 × 1.05	Multi-band 2D spin-echo EPI, factor of 2	1388	7T	AP and PA

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

3. We introduce a learning method that leverages T_{1w} images in the training of TS-Net. The motivation is that the T_{1w} image is widely considered as a *gold standard* representation of a subject's brain anatomy [17], and it is readily available in brain studies [18]. To make TS-Net more applicable for general use, the T_{1w} image is used *only* in training for network regularization, but not in the inference phase.
4. We provide an extensive evaluation of the proposed TS-Net on four large public datasets from the Human Connectome Project (HCP) [19]. First, an ablation study is conducted to analyze the effects of using different similarity measures to train TS-Net, the effects of various components in the TS-Net framework, and the effects of using a pre-trained TS-Net when training a new dataset. Second, TS-Net is compared with three state-of-the-art SAC methods, i.e. TOPUP [10], TISAC [15], and S-Net [16], in terms of correction accuracy and processing time.

The remainder of this paper is organized as follows. Section 2 describes the materials and the proposed method. Section 3 presents the experimental results and Section 4 discusses the proposed method and results. Finally, Section 5 summarizes our work.

2. Materials and Methods

In this section, Section 2.1 describes the EPI datasets used for experiments. Section 2.2 introduces the proposed TS-Net method. Section 2.3 presents the methods used for conducting experiments.

2.1. EPI datasets

To evaluate the SAC methods, we used four EPI datasets (fMRI-3T, DWI-3T, fMRI-7T, and DWI-7T), which are the unprocessed data of the *Subjects with 7T MR Session* from the public Human Connectome Project repository. The functional and diffusion MRI datasets were used to study functional connectivity of the human brain and reconstruct the complex axonal fiber architecture, respectively [20,21]. These four datasets were acquired using different acquisition sequences, imaging modalities, field strengths, resolutions, and image sizes; thus, the datasets are diverse in size and distortion property. Table 1 shows a summary of the four datasets. Note that the apparent diffusion coefficient map was not acquired in the DWI datasets. The b-values were 1000, 2000, and 3000 s/mm^2 for the DWI-3T dataset, and 1000 and 2000 s/mm^2 for the DWI-7T dataset.

2.2. Proposed TS-Net method

This section introduces a 3D anatomy-guided deep learning framework, called TS-Net, to correct the susceptibility artifacts in a 3D reversed-PE image pair (see Fig. 1). The proposed TS-Net includes a deep convolutional network to map the 3D image pair to the 3D displacement field $\mathbf{U}$. It also has a 3D spatial transform unit to unwarp the input-distorted images with the predicted displacement field, providing the corrected images. In contrast to existing SAC methods [15,16], TS-Net estimates the 3D displacement field, or three displacement values for each voxel. Thus, the displacement field $\mathbf{U}$ can be represented as $[U_x, U_y, U_z]$, where U_d is the displacement field in the d direction.

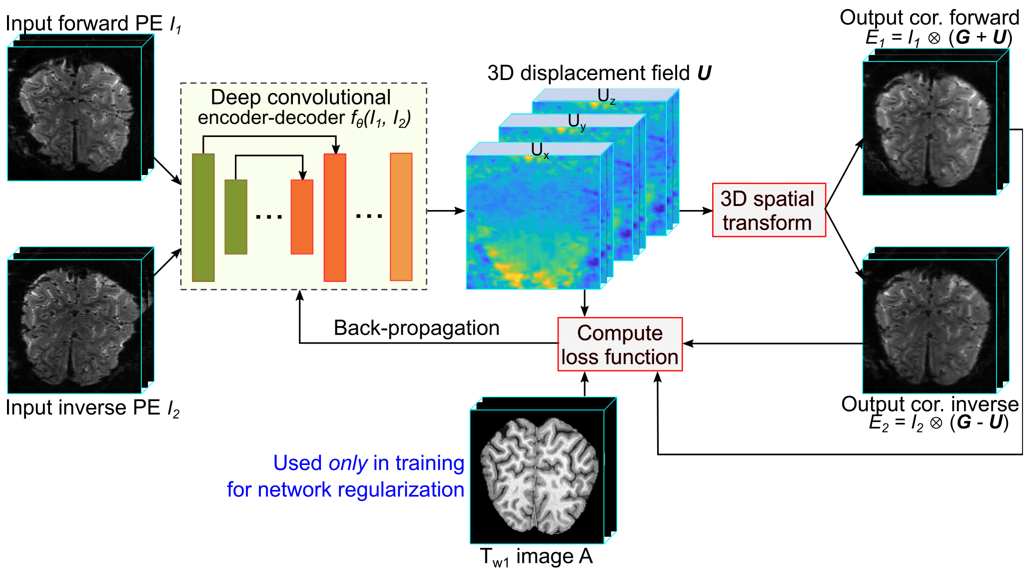


Figure 1. The proposed learning framework (TS-Net) for correcting the SAs in reversed-PE images. TS-Net accepts a pair of 3D reversed-PE images and produces the 3D displacement field and the corrected images.

The 3D spatial transform unit is the interpolation operator to unwarp or resample the input images by the estimate displacement field [22]. Let $\mathbf{U}$ denote the displacement field of image I_1 to the corrected image, then $-\mathbf{U}$ is the displacement field of image I_2 to the corrected image because of the inverse distortion property of the reversed-PE image pair. The spatial transform unit produces the corrected images, expressed as $E_1 = \left[I_1 \otimes (\mathbf{G} + \mathbf{U})\right]$, and $E_2 = \left[I_2 \otimes (\mathbf{G} - \mathbf{U})\right]$, where $\otimes$ is the linear interpolation and $\mathbf{G} = [G_x, G_y, G_z]$ is the regular grids in the x, y, and z directions.

The deep convolutional network can be considered as a mapping function $f_{\boldsymbol{\theta}} : (I_1, I_2) \rightarrow \mathbf{U}$, where $\boldsymbol{\theta}$ is the set of network parameters. The deep network, which is inspired by S-Net [16], U-Net [23], and DL-GP [24], is U-Net-like architecture with an encoder and a decoder (see Fig. 2). The encoder takes a two-channel input (which is the reverse PE image pair) and extracts the latent features. The decoder takes the latent features to predict the displacement field.

Both the encoder and the decoder use a kernel size of $3 \times 3 \times 3$ voxels for their convolutional layers to extract information from the neighboring voxels. This kernel size is selected because it requires fewer trainable parameters than larger kernel sizes, thereby improving computational efficiency. Each convolutional layer is followed by a BN layer to mitigate changes in the distribution of the convolutional layer's input [25].

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

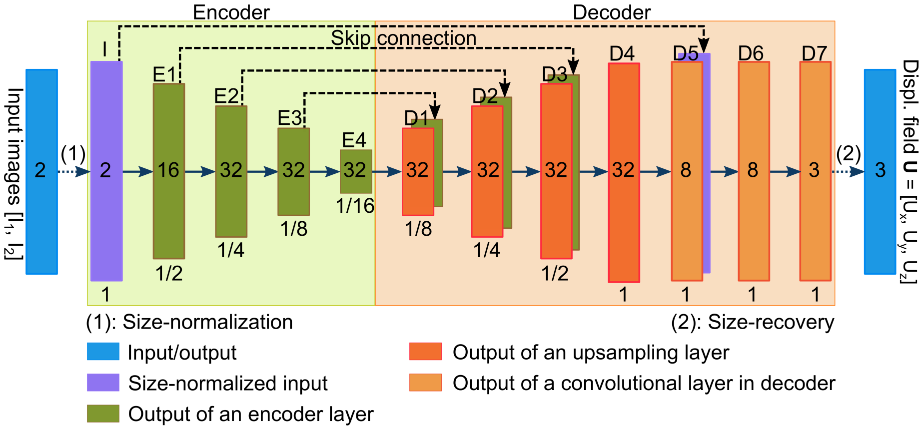


Figure 2. The convolutional encoder-decoder for mapping a pair of reversed-PE images to the 3D displacement field. *Box*: output feature maps of a layer. *Number inside each box*: number of feature maps in the layer. *Number below each box*: feature map size relative to the full input image size.

displacements in the EPI images are small and sensitive to image interpolation. Note that the configuration of the introduced convolutional encoder-decoder, e.g. the number of layers, batch normalization, and upsampling layers, was experimentally selected, see Section 3.1.

In our previous deep-learning-based SAC method [16], the network parameters $\boldsymbol{\theta}$ are estimated by optimizing the objective function that promotes the similarity between the pair of corrected images and enforces the local smoothness of the predicted displacement field. In this study, we regularize the training by introducing a T_{1w}-based regularizer to the loss function. This regularizer can improve the TS-Net training as the T_{1w} image is widely considered a gold standard representation of a subject's brain anatomy [17]. Note that T_{1w} images are used in the training phase, not in the testing phase.

The T_{1w}-based regularizer penalizes the distances from the corrected images to the corresponding T_{1w} structural image. Since T_{1w} and EPI are in different modalities, we use the normalized mutual information (NMI) to measure the similarity between the output images and the T_{1w} image because it is effective for multi-modal images. Let A denote the T_{1w} image, then the T_{1w}-based regularizer is defined as

$$\mathcal{L}_{\text{anat}}(E_1, E_2, A) = 1 - \frac{\text{NMI}(E_1, A) + \text{NMI}(E_2, A)}{2}. \tag{1}$$

The loss for TS-Net training is

$$\mathcal{L}(I_1, I_2, A, \mathbf{U}) = \mathcal{L}_{\text{sim}}(E_1, E_2) + \lambda\mathcal{L}_{\text{smooth}}(\mathbf{U}) + \gamma\mathcal{L}_{\text{anat}}(E_1, E_2, A), \tag{2}$$

where $\mathcal{L}_{\text{sim}}$ is the dissimilarity between the pair of corrected images. $\mathcal{L}_{\text{smooth}}$ is the diffusion regularizer, denoting the non-smoothness of the predicted displacement field. The positive and user-defined regularization parameters λ and γ represent the trade-off between the similarity of the corrected images, the smoothness of the displacement field, and the similarity of the T_{1w} image to the output images.

Since the corrected images E_1 and E_2 have the same modality, we investigate three possible unimodal similarity metrics: mean squared error (MSE), local cross-correlation (LCC) [26], and local normalized cross-correlation (LNCC) [27] (refer to Appendix (A) for a detailed description of the metrics). We experimentally found that LNCC metric is the best choice in terms of the trade-off between training accuracy and processing time (see the analysis in Section 3.1). Thus, LNCC is used as the $\mathcal{L}_{\text{sim}}$.

2.3. Experimental methods

To evaluate TS-Net, for each dataset, we first split the subjects randomly into two parts: A and B. Then, the training set was formed by randomly selecting reversed-PE image pairs of each subject in Part A; this strategy reduces the data repetition of subjects. The test set was formed from all reversed-PE pairs of each subject in Part B. The training sets were used to select the hyper-parameters and train the TS-Net models, and the test sets were used to evaluate the correction accuracy of the TS-Net models. The training set of each dataset was further divided into a training set and a validation set with a ratio of 9 : 1. Table 2 summarizes the training, validation, and test sets of the four datasets.

Table 2: A summary of the training, validation, and test sets for each of the four datasets.

Datasets	Training set		Validation set		Test set	
	No. subjects	No. pairs	No. subjects	No. pairs	No. subjects	No. pairs
fMRI-3T	140	1685	16	187	26	1395
DWI-3T	135	392	15	44	30	90
fMRI-7T	138	2890	15	322	31	1269
DWI-7T	133	140	15	15	30	60

The proposed TS-Net was implemented using Keras [28] deep learning library. For training TS-Net, the Adam optimizer was used with the learning rate $\alpha = 0.001$, and the exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.999$, as suggested by Kingma and Ba [29]. The Tree of Parzen Estimator algorithm was used to select suitable values for regularization parameters λ and γ [30–32].In training each dataset, we selected the maximum batch size that could fit into the available GPU memory to reduce the training time. The batch sizes and regularization parameters used in training TS-Net are shown in Table 3.

Table 3: Values of hyper-parameters in training TS-Net on the four datasets.

Params	fMRI-3T	DWI-3T	fMRI-7T	DWI-7T
λ	0.1771	0.002	0.9323	0.025
γ	0.01	0.01	0.01	0.01
Batch size	4	1	1	1

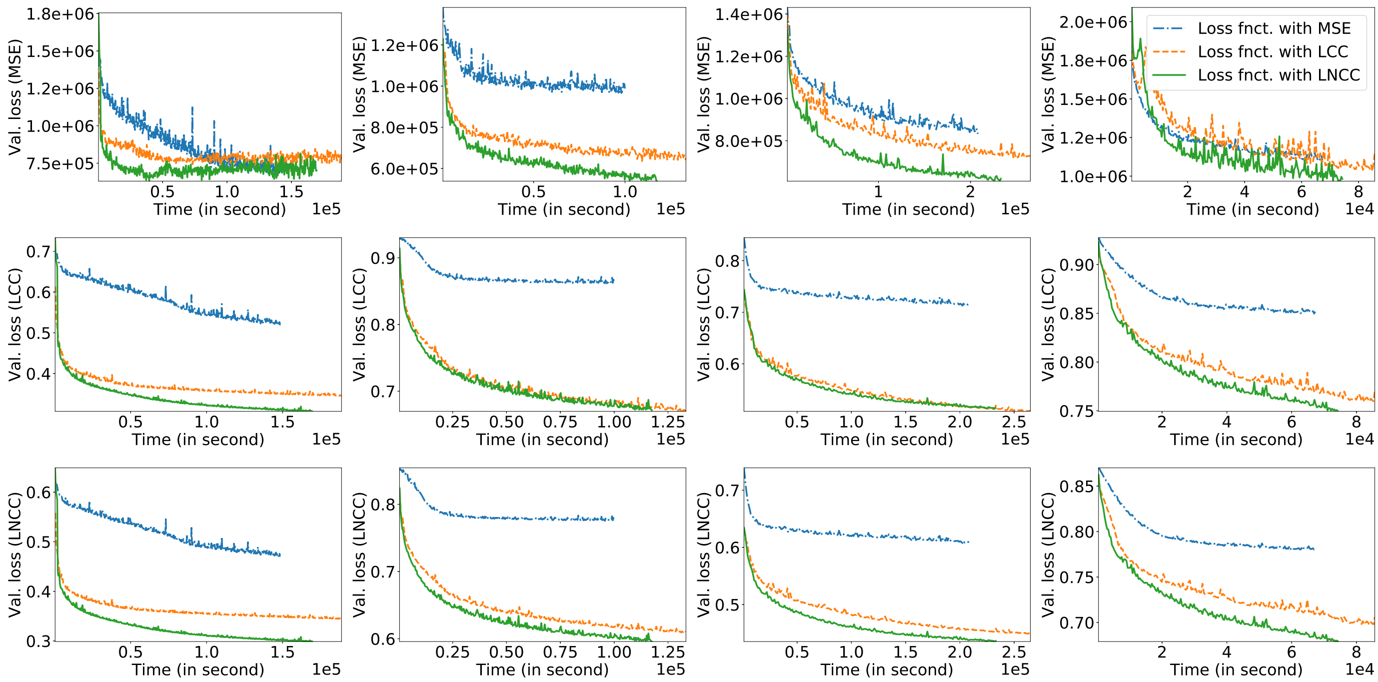
We then compared the proposed TS-Net with two iterative-optimization methods, i.e. TOPUP and TISAC, and a state-of-the-art deep learning method, i.e. S-Net. The comparison is in terms of the correction accuracy and processing speed. To evaluate the correction accuracy of the proposed method, we trained S-Net and TS-Net for 1500 epochs with each dataset. The trained models were used to compute the corrected image pairs of the test sets. For TOPUP[1] and TISAC, the corrected image pairs were obtained by implementing the iterative-optimization algorithms. Here, the correction accuracy is measured in terms of LNCC similarity between the pair of reversed-PE images.

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

3. Results

In this section, Section 3.1 presents results of the ablation study. Section 3.2 shows the results of the proposed method and other representative SAC methods in terms of correction accuracy and processing time.

[1] We used the TOPUP implementation in the FSL package, Website: fsl.fmrib.ox.ac.uk/fsl/fslwiki/topup



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

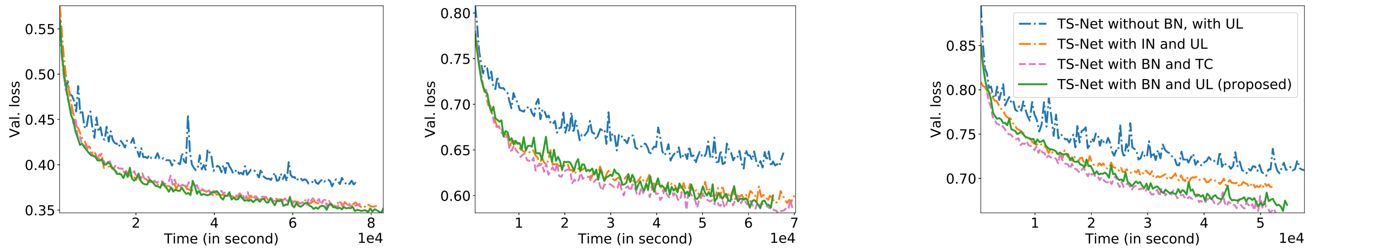
Figure 3. Validation loss of the models trained with three types of similarity loss (MSE, LCC, and LNCC) versus training time (in second) on the four datasets: (a) fMRI-3T; (b) DWI-3T; (c) fMRI-7T; and (d) DWI-7T. *Top row*: validation loss in terms of MSE. *Middle row*: validation loss in terms of LCC.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

3.1. Ablation study of the proposed method

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

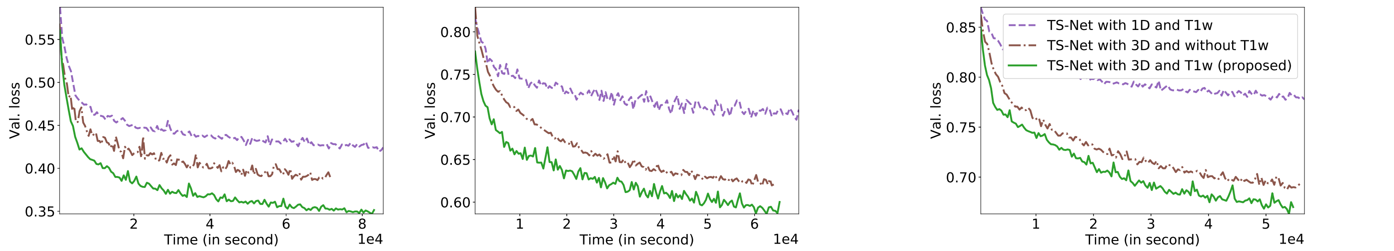
Effects of similarity measures in network training: In this experiment, for each training set, we trained TS-Net models using different similarity losses: (i) MSE; (ii) LCC; and (iii) LNCC. The effects of using different similarity measures were evaluated in two aspects: the validation loss and the training time of each epoch. The validation loss was measured as the mean similarity measures for output image pairs across subsets of the training sets. We conducted the experiments on the four datasets: fMRI-3T, DWI-3T, fMRI-7T, and DWI-7T. Fig. 3 shows the validation loss versus time when training TS-Net with the similarity loss as MSE, LCC, and LNCC. It can be seen that TS-Net trained with the LNCC measure produces the lowest validation loss, while TS-Net trained with the MSE measure produces the highest validation loss. TS-Nets trained with the LNCC and LCC measures produce a competitive LCC validation loss on two datasets (DWI-3T and fMRI-7T). Considering the validation loss versus the training time, it is clear that the LNCC measure is a better choice than the MSE and the LCC for training TS-Net. Based on this experiment, the LNCC metric was subsequently used as the similarity loss for all the remaining experiments.

Effects of the network configurations in TS-Net: In this experiment, we analyzed the effects of four different network configurations: (i) TS-Net without batch normalization and with upsampling layer (UL) (ii) TS-Net with instance normalization (IN) [33], and with UL; (iii) TS-Net with BN and transposed convolution (TC) [34]; and (iv) TS-Net with BN and UL (proposed method). The validation loss during the training phase was computed as the average LNCC measure between the output image pairs, across subsets



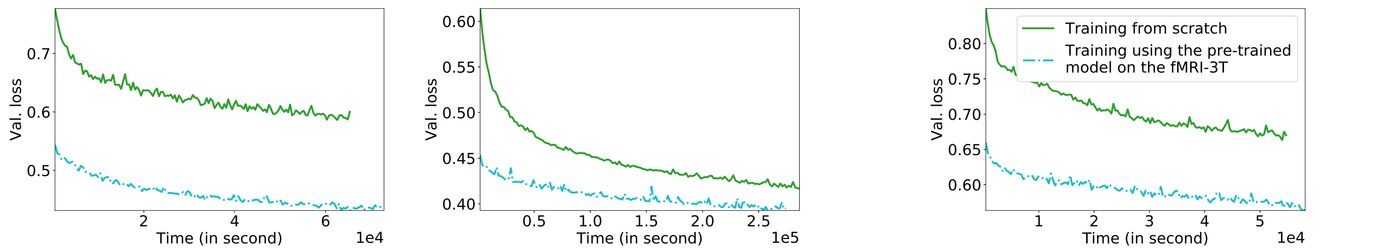
fMRI-3T DWI-3T DWI-7T

(a) Comparison of the validation loss on four models: (i) TS-Net without batch normalization and with upsampling layer (UL); (ii) TS-Net with instance normalization and UL; (iii) TS-Net with batch normalization (BN) and transposed convolution; and (iv) TS-Net with BN and UL (proposed method).



fMRI-3T DWI-3T DWI-7T

(b) Comparison of the validation loss on three models: (i) TS-Net with 1D distortion model and T_{1w} guidance; (ii) TS-Net with 3D distortion model and without T_{1w} guidance; and (iii) TS-Net with 3D distortion model and T_{1w} guidance (proposed method).



DWI-3T fMRI-7T and DWI-7T

(c) Comparison of the validation loss on two models trained: (i) from scratch; and (ii) using the pre-trained model of the fMRI-3T dataset.

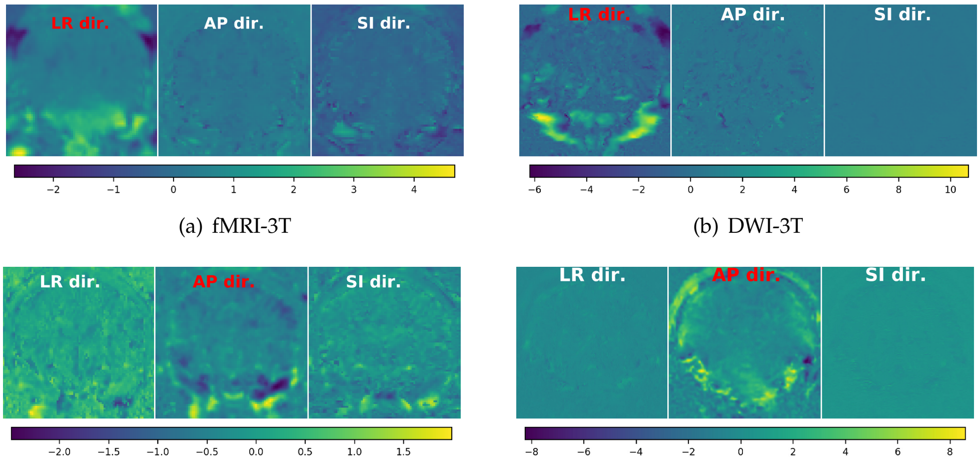
Figure 4. Ablation study of TS-Net in terms of: (a) network configurations; (b) 3D distortion model and anatomical guidance; and (c) using a pre-trained model. Plots show the validation loss of trained models versus training time (in second).

of the training sets. This validation loss was then used to compare different network configurations.

Fig. 4(a) shows the validation loss versus the training time on three datasets: fMRI-3T, DWI-3T, and DWI-7T; each subfigure includes the validation loss for the four network configurations. Several observations can be made. First, using batch normalization (proposed TS-Net, green curve) provides a lower validation loss compared to not using batch normalization (blue curve). Second, using batch normalization (proposed TS-Net, green curve) provides a similar or lower validation loss compared to using instance normalization (orange curve). Third, using the upsampling layer (proposed TS-Net, green curve) has a similar validation loss compared to using the transpose convolution (magenta curve). These results justify our selected configuration for TS-Net.

Effects of using the 3D distortion model and anatomical guidance by T_{1w}: In this experiment, we trained three types of networks: (i) TS-Net with the 1D distortion model as used in S-Net [16]; (ii) TS-Net with 3D distortion model and without T_{1w} guidance; and (iii) TS-Net with the 3D distortion model and T_{1w} guidance (proposed method). Fig. 4(b) shows the validation loss versus the training time on three datasets: fMRI-3T, DWI-3T, and DWI-7T. Several observations can be made. First, the proposed TS-Net with

T_{1w} guidance (green-solid curve) has lower validation losses than the TS-Net without T_{1w} guidance (brown dash-dotted curve). This result shows that incorporating T_{1w} guidance can improve the correction accuracy. Second, the proposed TS-Net using the 3D distortion model (green-solid curve) produces significantly lower validation losses than TS-Net using the 1D distortion model (magenta-dashed curve). This result shows that the 3D distortion model used in the proposed TS-Net provides more accurate correction than the 1D distortion model (i.e. only along the phase-encoding direction), which is used in S-Net and existing iterative-optimization SAC methods.



(a) fMRI-3T (b) DWI-3T

(c) fMRI-7T (d) DWI-7T

Figure 5. Samples of three predicted displacement fields (in voxel) of TS-Net from the four test sets. In each subfigure, *left image*: displacement field in the left-right (LR) direction; *middle image*: displacement field in the anterior-posterior (AP) direction; and *right image* displacement field in the superior-inferior (SI) direction. The dominant phase-encoding dimension (direction) is shown in red text; the other two other dimensions are shown in white text.

Effects of using a pre-trained TS-Net: In this experiment, we explored whether using a TS-Net model pre-trained on one dataset can reduce the training time on another dataset, compared to a randomly initialized TS-Net. To this end, we trained two TS-Net models: (i) from scratch; and (ii) using an *initial* network, which had been pre-trained for 1500 epochs on the fMRI-3T dataset. Fig. 4(c) shows the validation loss versus training time on three datasets: DWI-3T, fMRI-7T, and DWI-7T. The figure shows that the validation loss when training TS-Net using a pre-trained model (cyan dash-dotted curve) is much lower than when training from scratch (green-solid curve). The result suggests that TS-Net is able to learn generalized features for correcting the susceptibility artifacts from one dataset. Subsequently, adopting the learned features in training other datasets leads to a faster converge.

Visualization of the predicted displacement fields: Fig. 5 shows the samples of the displacement field estimated by the trained TS-Net for the four test sets. The displacement field is shown in three directions (left-right, anterior-posterior, and superior-inferior). TS-Net can estimate the geometric distortions along the directions that are not the dominant PE direction. The visual results indicate that TS-Net is able to predict realistic 3D displacement fields, i.e. the displacements in the phase-encoding direction are dominant than the one on the other two directions.

3.2. Comparison with other methods

This section compares TS-Net with three SAC methods, i.e. TOPUP, TISAC, and S-Net. Fig. 6 shows sample slices of uncorrected and corrected images from each of the four test sets. Each example includes two reversed-PE images (Rows 1 and 2) and the absolute

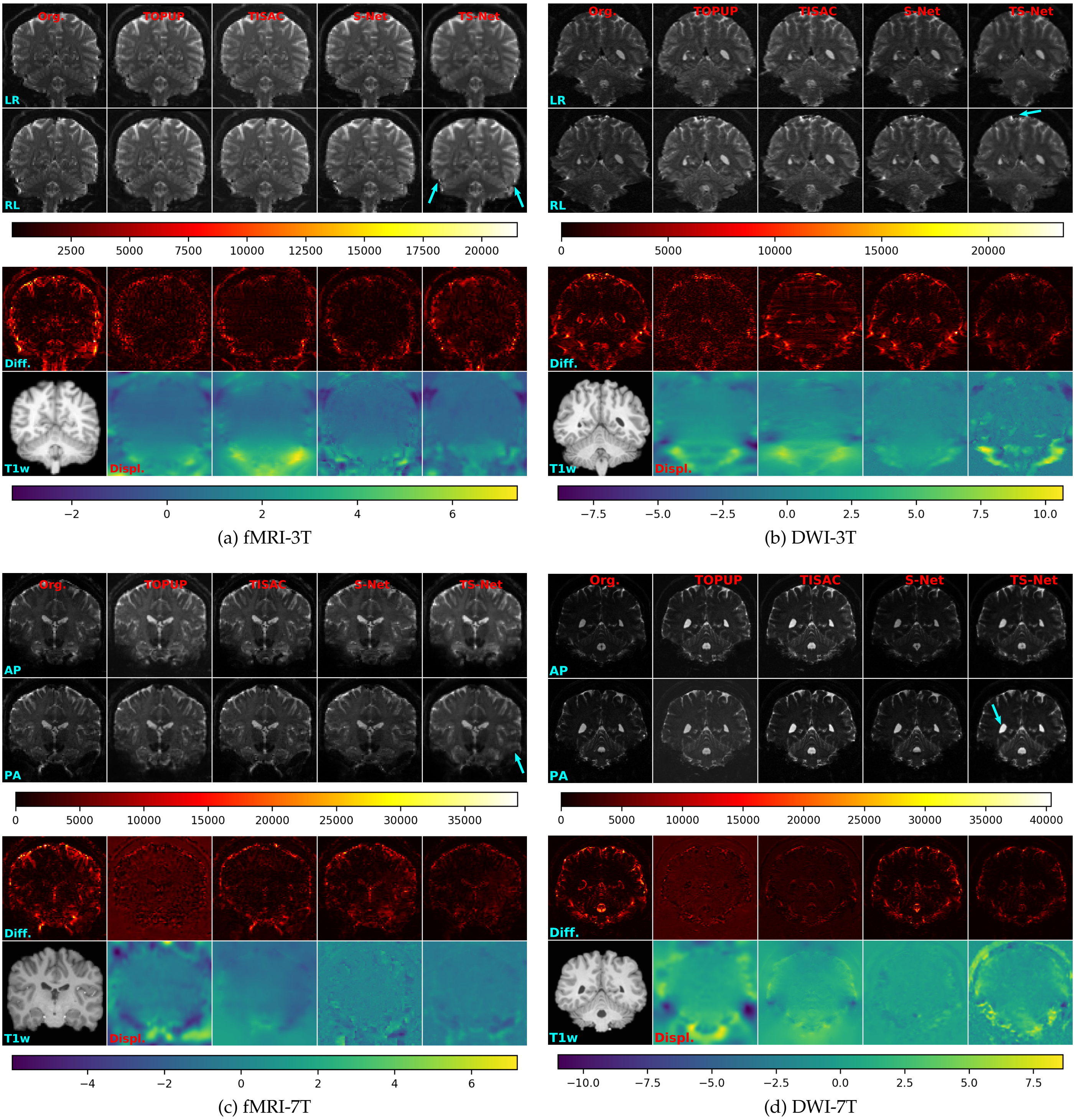


Figure 6. Sample visual results of SAC methods from the four test sets. In each subfigure, *Column 1*: input uncorrected images. *Columns 2, 3, 4, and 5*: output corrected images produced by TOPUP, TISAC, S-Net, and TS-Net, respectively. *Rows 1 and 2*: reversed phase-encoding EPI images. *Row 3*: the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

difference between the two images (Row 3). The arrows indicate the regions where TS-Net produces significantly improved correction in comparison with three other SAC methods. It can be seen that TS-Net removes distortions in the uncorrected images significantly. In general, TS-Net produces the output images that are comparable to or better than the outputs of TOPUP, TISAC, and S-Net. Note that the SAC methods work with 3D images; however, for visualization, 2D slices are presented in the figures. For a larger view of the TS-Net outputs, see Fig. A1 in Appendix (B).

Table 4: Accuracy in terms of local normalized cross-correlation for different test sets: fMRI-3T, DWI-3T, fMRI-7T, DWI-7T.

Datatypes	fMRI-3T	DWI-3T	fMRI-7T	DWI-7T
	mean ± std	mean ± std	mean ± std	mean ± std
Uncorrected	0.335* ± 0.023	0.142* ± 0.020	0.229* ± 0.023	0.120* ± 0.018
TOPUP	**0.753*** ± 0.024	0.468* ± 0.031	0.583* ± 0.024	0.371* ± 0.025
TISAC	0.674* ± 0.036	0.436* ± 0.058	0.427* ± 0.036	0.364* ± 0.048
S-Net	0.608* ± 0.027	0.242* ± 0.039	0.412* ± 0.027	0.182* ± 0.025
TS-Net	0.692 ± 0.022	**0.571** ± 0.034	**0.648** ± 0.022	**0.398** ± 0.031

The asterisk symbol (*) indicates that the computed P is less than 0.001 for the null hypothesis $\mathcal{H}_0 : m_{\text{TS-Net}} = m_{\text{other}}$. A P value below 0.001 means that the null hypothesis is rejected at a confidence level of 99.9%. In other words, the similarity measure LNCC of TS-Net is significantly different from the compared method.

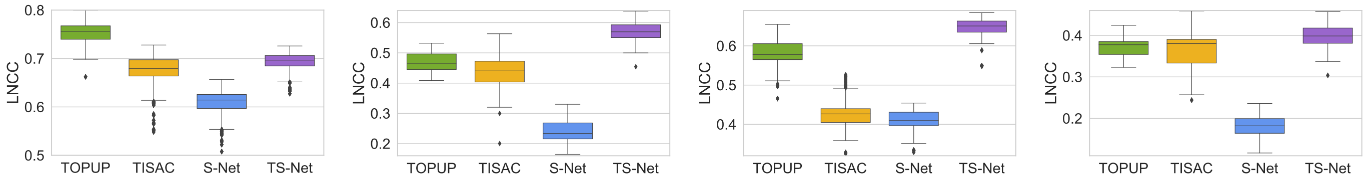
Table 4 summarizes the accuracy of uncorrected and corrected images in terms of LNCC on four different test sets. Paired t-tests were performed on the LNCC measures between TS-Net outputs and each of four image types: uncorrected images, TOPUP outputs, TISAC outputs, and S-Net outputs. The null hypothesis is $\mathcal{H}_0 : m_{\text{S-Net}} = m_{\text{other}}$. All computed P values are smaller than 0.001; this indicates that the null hypothesis is rejected at a confidence level of 99.9%. In other words, TS-Net produces image pairs with significant differences (i.e. improvements) in terms of accuracy compared to the output image pairs of other methods.

Table 5: Processing time (in second) of SAC methods for correcting a pair of reversed-PE images.

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

For visual clarity, Fig. 7 shows the box plots for comparing the LNCC measures of the four SAC methods. The results in Table 4 and Fig. 7 show three notable observations. First, TS-Net produces output images that have significantly higher LNCC measures than the uncorrected images; in other words, TS-Net does reduce the susceptibility artifacts. Second, TS-Net produces output images that have higher LNCC measures than the outputs of the TISAC method in 4 out of 4 datasets, and the outputs of the TOPUP methods in 3 out of 4 datasets. This means that TS-Net has better correction accuracy compared to the two iterative-optimization methods, i.e. TISAC and TOPUP. Third, TS-Net also produces higher LNCC measures than S-Net in 4 out of 4 datasets. Compared to S-Net, the proposed TS-Net has several differences, one of which is its use of T_{1w} images in training. This result demonstrates that including the *gold-standard* representation of a subject's brain anatomy helps regularize the susceptibility artifact correction in TS-Net. Note that TS-Net does not require the T_{1w} image in the inference phase, which explains its comparable processing speed with S-Net, as analyzed next.

To compare the processing speed, we first randomly selected 50 distorted image pairs for each of the four datasets. We then recorded the time for correcting the selected image pairs by four SAC methods: TOPUP, TISAC, S-Net, and TS-Net. Table 5 shows the



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

Figure 7. Comparisons of the proposed TS-Net versus other three SAC methods in terms of the LNCC-based accuracy on the test sets. Due to differences in the LNCC ranges of the datasets, the plots are drawn in different y-axis ranges for clarity. In each box plot, the *top line* is the maximum LNCC value excluding the outliers; the *bottom line* is the minimum LNCC value excluding the outliers; the *middle line* is the median LNCC value; the *solid rectangle* is the interquartile range of the LNCC values; and the *points* are the outliers.

average processing time per image pair of TS-Net and the three SAC methods. Over the four datasets, TS-Net is 396.72 times faster than TOPUP, 29.45 times faster than TISAC, and only 1.05 times slower than S-Net. Both deep learning-based SAC methods (TS-Net and S-Net) can be accelerated by five times when using the GPU instead of the CPU. Note that in the experiments for all datasets, the proposed TS-Net has 260,187 trainable parameters, whereas the S-Net model has 259,241 trainable parameters. In other words, the proposed TS-Net requires only 0.36% more trainable parameters than S-Net.

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

4. Discussion

This section discusses the proposed TS-Net in three aspects: robustness, generalizability, and feasibility. In terms of robustness, TS-Net can predict realistic 3D displacement fields, i.e. the most dominant displacements in the phase-encoding direction regardless of the PE direction order, resulting in high-quality corrected images. The experiments conducted on four different datasets show that TS-Net performed consistently on different image resolutions, image sizes, image modalities, and training set sizes. Furthermore, it can cope with different phase-encoding directions.

In terms of generalizability, TS-Net is able to learn the generalized features of the susceptibility artifacts in reversed-PE image pairs from one dataset. A trained TS-Net can be easily transferred to a new dataset, effectively reducing the training time. This observation is similar to the generalization capability of the deep networks [35]. Therefore, TS-Net can employ the network initialization techniques, e.g. MAML [36] and Reptile [37], to address the problem of long training time, which is a common bottleneck in deep learning algorithms.

In terms of feasibility, TS-Net can produce higher accuracy than the state-of-the-art SAC methods, while having fast processing time. To correct a pair of distorted images, TS-Net only takes less than 5 seconds using CPU or less than 1 second using GPU. These high-accuracy and high-speed capabilities allow TS-Net to be applied in many applications. For example, the TS-Net can be integrated into the MRI scanner to correct SAs in real-time; this is typically not possible with the traditional reversed-PE SAC methods because they are slow.

5. Conclusions

This paper presented an end-to-end 3D anatomy-guided deep learning framework, TS-Net, to correct the susceptibility artifacts in reversed phase-encoding 3D EPI image pairs. The proposed TS-Net contains a deep convolutional network to predict the displacement field in all three directions. The corrected images are then generated by feeding the predicted displacement field and input images into a 3D spatial transform unit. In the training phase, the proposed TS-Net additionally utilizes T_{1w} images to regularize the susceptibility artifact correction. However, the T_{1w} images are not used in the inference phase to simplify the use of TS-Net.

The visual analysis shows that TS-Net is able to estimate the realistic 3D displacement field, i.e. the displacements are dominant in the phase-encoding direction than the other two directions. Evaluation on the four large datasets also demonstrates that the proposed TS-Net provides higher correction accuracy than TISAC and S-Net in all four datasets, and TOPUP in three out of four datasets. Over the four datasets, TS-Net runs significantly faster than the iterative-optimization SAC methods: 396.72 times faster than TOPUP and 29.45 times faster than TISAC. TS-Net is slightly slower than S-Net, but it still meets the real-time correction requirement of MRI scanners. Furthermore, the training time of TS-Net on a new dataset can be reduced by using a pre-trained model.

Funding: This research was funded by Discovery Projects (DP170101778 and DP190100607) from the Australian Research Council and a Matching scholarship from the University of Wollongong.

Acknowledgments: This research used the data provided by the Human Connectome Project.

Conflicts of Interest: The authors have declared that no competing interests exist.

Appendix A. Similarity metrics

This section presents the three similarity metrics, i.e. MSE, LCC, and LNCC, which are used in $\mathcal{L}_{\text{sim}}$.

Appendix A.1. Mean squared error

The **MSE** between two images E_1 and E_2 is defined as

$$\text{MSE}(E_1,\, E_2) = \frac{1}{|\Omega|} \sum_{\mathbf{p} \in \Omega} \left[E_1(\mathbf{p}) \,-\, E(\mathbf{p}) \right]^2, \tag{A1}$$

where $\Omega \in \mathcal{R}^3$ is the image domain and $|\Omega|$ is the total number of image indexes. A smaller value of MSE indicates a higher similarity between the images. Thus, the $\mathcal{L}_{\text{sim}}$ loss based on the MSE measure is

$$\mathcal{L}_{\text{sim}}^{\text{MSE}}(E_1,\, E_2) = \text{MSE}(E_1,\, E_2). \tag{A2}$$

Appendix A.2. Local cross-correlation

The **LCC** 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 centered image $\hat{X}$ is computed as

$$\hat{X} = X - \bar{X}. \tag{A3}$$

For a given voxel $\mathbf{p} = (x, y, z)$, let $W(\mathbf{p})$ denote the set of voxels in the $n \times n \times n$ cube centered on $\mathbf{p}$. For a pair of images E_1 and E_2, we compute a local correlation coefficient image C:

$$C(\mathbf{p}) = \frac{\left(\sum_{\mathbf{p}_i \in W(\mathbf{p})} [\hat{E}_1(\mathbf{p}_i)\; \hat{E}_2(\mathbf{p}_i)] \right)^2}{\sum_{\mathbf{p}_i \in W(\mathbf{p})} [\hat{E}_1(\mathbf{p}_i)]^2 \; \sum_{\mathbf{p}_i \in W(\mathbf{p})} [\hat{E}_2(\mathbf{p}_i)]^2}. \tag{A4}$$

The LCC measure for images E_1 and E_2 is now defined as the mean intensity of the local correlation image C:

$$\text{LCC}(E_1,\, E_2) = \frac{1}{|\Omega|} \sum_{\mathbf{p} \in \Omega} C(\mathbf{p}). \tag{A5}$$

A higher LCC indicates more similarity between two output images. We now can express the $\mathcal{L}_{\text{sim}}$ loss based on the LCC measure as

$$\mathcal{L}_{\text{sim}}^{\text{LCC}}(E_1, E_2) = 1 - \text{LCC}(E_1, E_2). \tag{A6}$$

380 *Appendix A.3. Local normalized cross-correlation*

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

$$\tilde{X}(\mathbf{p}) = \sum_{\mathbf{p}_i \in W(\mathbf{p})} [X(\mathbf{p}_i)]^2 - \frac{1}{n^3}\Big[\sum_{\mathbf{p}_i \in W(\mathbf{p})} X(\mathbf{p}_i)\Big]^2. \tag{A7}$$

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

$$R(\mathbf{p}) = \sum_{\mathbf{p}_i \in W(\mathbf{p})} [E_1(\mathbf{p}_i)\, E_2(\mathbf{p}_i)] - \frac{1}{n^3} \sum_{\mathbf{p}_i \in W(\mathbf{p})} E_1(\mathbf{p}_i) \sum_{\mathbf{p}_i \in W(\mathbf{p})} E_2(\mathbf{p}_i). \tag{A8}$$

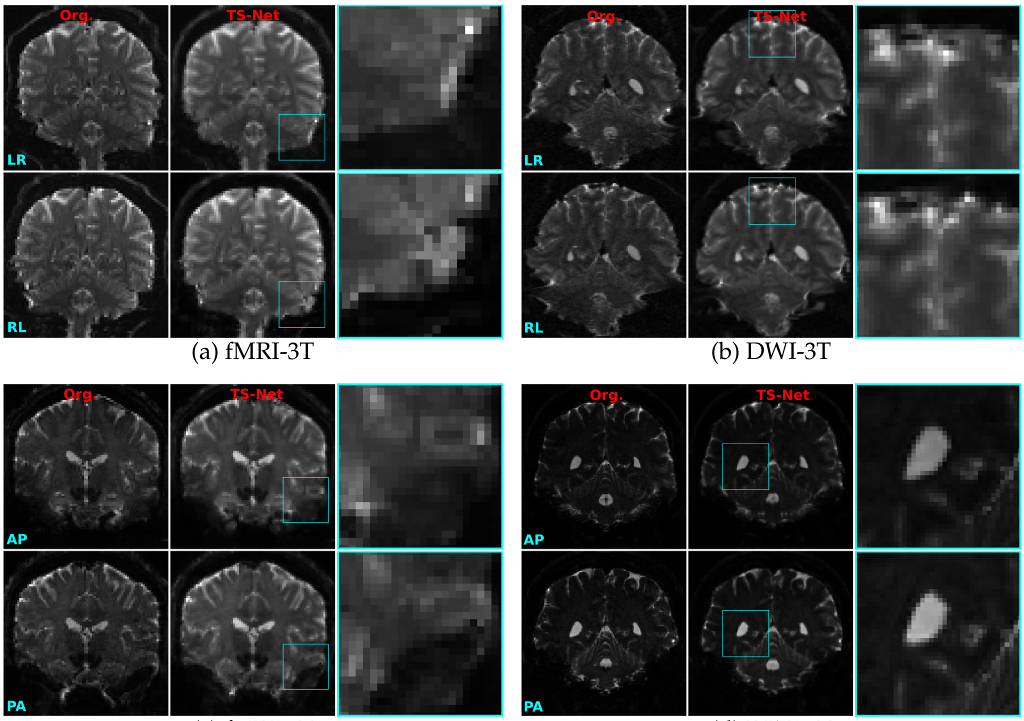
The LNCC between two images E_1 and E_2 is given by

$$\text{LNCC}(E_1, E_2) = \frac{1}{|\Omega|} \sum_{\mathbf{p} \in \Omega} \frac{[R(\mathbf{p})]^2}{\tilde{E}_1(\mathbf{p})\, \tilde{E}_2(\mathbf{p})}. \tag{A9}$$

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

$$\mathcal{L}_{\text{sim}}^{\text{LNCC}}(E_1, E_2) = 1 - \text{LNCC}(E_1, E_2). \tag{A10}$$

381 Appendix B. Supplementary data



(a) fMRI-3T (b) DWI-3T

(c) fMRI-7T (d) DWI-7T

Figure A1. Larger view of the TS-Net outputs from the four test sets. In each subfigure,*Column 1*: input uncorrected images. *Columns 2*: output corrected images produced by TS-Net. *Columns 3*: the zoomed view of cyan rectangles from the TS-Net output.

References

1. Poustchi-Amin, M.; Mirowitz, S.A.; Brown, J.J.; McKinstry, R.C.; Li, T. Principles and applications of echo-planar imaging: a review for the general radiologist. *Radiographics* **2001**, *21*, 767–779.
2. Matthews, P.M.; Honey, G.D.; Bullmore, E.T. Applications of fMRI in translational medicine and clinical practice. *Nat. Rev. Neurosci.* **2006**, *7*, 732–744.
3. Baars, B.J.; Gage, N.M., Brain imaging. In *Fundamentals of Cognitive Neuroscience*; Academic Press: San Diego, 2013; book section 5, pp. 109–140.
4. Chang, H.; Fitzpatrick, J.M. A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. *IEEE Trans. Image Process.* **1992**, *11*, 11.
5. Schmitt, F., Echo-Planar Imaging. In *Brain Mapping - An Encyclopedic Reference*; Academic Press, 2015; Vol. 1, book section 6, pp. 53–74.
6. Chan, R.W.; von Deuster, C.; Giese, D.; Stoeck, C.T.; Harmer, J.; Aitken, A.P.; Atkinson, D.; Kozerke, S. Characterization and correction of Eddy-current artifacts in unipolar and bipolar diffusion sequences using magnetic field monitoring. *J. Magn. Reson.* **2014**, *244*, 74–84. doi:10.1016/j.jmr.2014.04.018.
7. Irfanoglu, M.O.; Sarlls, J.; Nayak, A.; Pierpaoli, C. Evaluating corrections for Eddy-currents and other EPI distortions in diffusion MRI: methodology and a dataset for benchmarking. *Magn. Reson. Med.* **2019**, *81*, 2774–2787. doi:10.1002/mrm.27577.
8. Jezzard, P.; Balaban, R.S. Correction for geometric distortion in echo planar images from B0 field variations. *Magn. Reson. Med.* **1995**, *34*, 65–73.
9. Holland, D.; Kuperman, J.M.; Dale, A.M. Efficient correction of inhomogeneous static magnetic field-induced distortion in echo planar imaging. *NeuroImage* **2010**, *50*, 175–184.
10. Andersson, J.L.R.; Skare, S.; Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* **2003**, *20*, 870–888.
11. Ruthotto, L.; Kugel, H.; Olesch, J.; Fischer, B.; Modersitzki, J.; Burger, M.; Wolters, C.H. Diffeomorphic susceptibility artifact correction of diffusion-weighted magnetic resonance images. *Phys. Med. Biol.* **2012**, *57*, 5715–5731.
12. Hedouin, R.; Commowick, O.; Bannier, E.; Scherrer, B.; Taquet, M.; Warfield, S.K.; Barillot, C. Block-matching distortion correction of echo-planar images with opposite phase encoding directions. *IEEE Trans. Med. Imaging* **2017**, *36*, 1106–1115.
13. Irfanoglu, M.O.; Modia, P.; Nayaka, A.; Hutchinson, E.B.; Sarllsc, J.; Pierpaoli, C. DR-BUDDI (diffeomorphic registration for blip-up blip-down diffusion imaging) method for correcting echo planar imaging distortions. *NeuroImage* **2015**, *106*, 284–299.
14. Duong, S.T.M.; Schira, M.M.; Phung, S.L.; Bouzerdoum, A.; Taylor, H.G.B. Anatomy-guided inverse-gradient susceptibility artefact correction method for high-resolution fMRI. Proc. IEEE Int. Conf. Acoust. Speech Signal Process., 2018, pp. 786–790.
15. Duong, S.T.M.; Phung, S.L.; Bouzerdoum, A.; Taylor, H.G.B.; Puckett, A.M.; Schira, M.M. Susceptibility artifact correction for sub-millimeter fMRI using inverse phase encoding registration and T1 weighted regularization. *J. Neurosci. Methods* **2020**, *336*, 108625.
16. Duong, S.T.M.; Phung, S.L.; Bouzerdoum, A.; Schira, M.M. An unsupervised deep learning technique for susceptibility artifact correction in reversed phase-encoding EPI mages. *Magn. Reson. Imaging* **2020**, *71*, 1–10.
17. Howarth, C.; Hutton, C.; Deichmann, R. Improvement of the image quality of T1-weighted anatomical brain scans. *NeuroImage* **2006**, *29*, 930–937.
18. Polimeni, J.R.; Renvall, V.; Zaretskaya, N.; Fischl, B. Analysis strategies for high-resolution UHF-fMRI data. *NeuroImage* **2018**, *168*, 296–320.
19. Essen, D.C.V.; Ugurbil, K.; Auerbach, E.; Barch, D.; Behrens, T.E.J.; Bucholz, R.; Chang, A.; Chen, L.; Corbetta, M.; Curtiss, S.W.; Penna, S.D.; Feinberg, D.; Glasser, M.F.; Harel, N.; Heath, A.C.; Larson-Prior, L.; Marcus, D.; Michalareas, G.; Moeller, S.; Oostenveld, R.; Petersen, S.E.; Prior, F.; Schlaggar, B.L.; Smith, S.M.; Snyder, A.Z.; Xu, J.; Yacoub, E. The human connectome project: a data acquisition perspective. *NeuroImage* **2012**, *62*, 2222–2231.
20. Essen, D.C.V.; Smith, S.M.; Barch, D.M.; Behrens, T.E.J.; Yacoub, E.; Ugurbil, K. The WU-Minn human connectome project: an overview. *NeuroImage* **2013**, *80*, 62–79.
21. Ugurbil, K.; Xu, J.; Auerbach, E.J.; Moeller, S.; Vu, A.T.; Duarte-Carvajalino, J.M.; Lenglet, C.; Wu, X.; Schmitter, S.; Moortele, P.F.V.d.; Strupp, J.; Sapiro, G.; Martino, F.D.; Wang, D.; Harel, N.; Garwood, M.; Chen, L.; Feinberg, D.A.; Smith, S.M.; Miller, K.L.; Sotiropoulos, S.N.; Jbabdi, S.; Andersson, J.L.R.; Behrens, T.E.J.; Glasser, M.F.; Essen, D.C.V.; Yacoub, E. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage* **2013**, *80*, 80–104.
22. Balakrishnan, G.; Zhao, A.; Sabuncu, M.R.; Guttag, J.; Dalca, A.V. VoxelMorph: a learning framework for deformable medical image registration. *IEEE Trans. Med. Imaging* **2019**, *38*, 1788–1800.
23. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: convolutional networks for biomedical image segmentation. Proc. Med. Image. Comput. Assist. Interv., 2015, pp. 234–241.
24. Nguyen, T.N.A.; Phung, S.L.; Bouzerdoum, A. Hybrid deep learning-Gaussian process network for pedestrian lane detection in unstructured scenes. *IEEE Trans. Neural Netw. Learn. Sys.* **2020**, pp. 1–15.
25. Ioffe, S.; Szegedy, C. Batch normalization: accelerating deep network training by reducing internal covariate shift. Proc. Int. Conf. Machine Learning, 2015, pp. 448–456.
26. Avants, B.B.; Epstein, C.L.; Grossman, M.; Gee, J.C. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med. Image. Anal.* **2008**, *12*, 26–41.

27. Baig, A.; Chaudhry, M.A.; Mahmood, A. Local normalized cross correlation for geo-registration. Proc. Int. Bhurban Conf. Appl. Sci. Technol., 2012, pp. 70–74.
28. Chollet, F. Keras. Available online: https://github.com/fchollet/keras, Accessed on: Apr. 1, 2020.
29. Kingma, D.; Ba, J. Adam: a method for stochastic optimization. *arXiv preprint* **2014**, p. arXiv:1412.6980.
30. Bergstra, J.; Bardenet, R.; Bengio, Y.; Kegl, B. Algorithms for hyper-parameter optimization. Proc. Int. Conf. Neural Inf. Process. Sys., 2011, pp. 2546–2554.
31. Bergstra, J.; Yamins, D.; Cox, D.D. Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. Proc. Int. Conf. Machine Learning, 2013, pp. I–115–I–123.
32. Bergstra, J.; Komer, B.; Eliasmith, C.; Yamins, D.; Cox, D.D. Hyperopt: a Python library for model selection and hyperparameter optimization. *Comput. Sci. Discov.* **2015**, *8*, 014008.
33. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance normalization: the missing ingredient for fast stylization. *arXiv preprint* **2016**, p. arXiv:1607.08022.
34. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., 2015, pp. 3431–3440.
35. Neyshabur, B.; Bhojanapalli, S.; McAllester, D.; Srebro, N. Exploring generalization in deep learning. Proc. Int. Conf. Neural Inf. Process. Sys., 2017, p. 5949–5958.
36. Finn, C.; Abbeel, P.; Levine, S. Model-agnostic meta-learning for fast adaptation of deep networks. Proc. Int. Conf. Machine Learning, 2017, Vol. 70, pp. 1126–1135.
37. Nichol, A.; Achiam, J.; Schulman, J. On first-order meta-learning algorithms. *arXiv preprint* **2018**, p. arXiv:1803.02999v3.