SACNet: A Multiscale Diffeomorphic Convolutional Registration Network with Prior Neuroanatomical Constraints for Flexible Susceptibility Artifact Correction in Echo Planar Imaging

Zilong Zeng[1,2,3], Jiaying Zhang[1,2,3], Xinyuan Liang[1,2,3], Lianglong Sun[1,2,3], Yihe Zhang[4], Weiwei Men[5,6], Yanpei Wang[1], Rui Chen[1], Haibo Zhang[1], Shuping Tan[7], Jia-Hong Gao[5,6,8], Shaozheng Qin[1,2,3,9], Qiqi Tong[10], Hongjian He[11,12], Sha Tao[1], Qi Dong[1], Yong He[1,2,3,9], Tengda Zhao[1,2,3*]

[1] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China

[2] Beijing Key Laboratory of Brain Imaging and Connectomics, Beijing Normal University, Beijing 100875, China

[3] IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China

[4] School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China

[5] Center for MRI Research, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China

[6] Beijing City Key Laboratory for Medical Physics and Engineering, Institute of Heavy Ion Physics, School of Physics, Peking University, Beijing 100871, China

[7] Beijing Huilongguan Hospital, Peking University Huilongguan Clinical Medical School, Beijing 100096, China

[8] IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China

[9] Chinese Institute for Brain Research, Beijing 102206, China

[10] Research Center for Healthcare Data Science, Zhejiang Lab, Hangzhou, Zhejiang, 311121, China

[11] Center for Brain Imaging Science and Technology, College of Biomedical Engineering and Instrumental Science, Zhejiang University, Hangzhou, Zhejiang, 310058, China

[12] School of Physics, Zhejiang University, Hangzhou, Zhejiang, 310058, China

*** Corresponding author:**

Tengda Zhao, Ph.D., Email: tengdazhao@bnu.edu.cn

Manuscript information: 29 text pages, 6 figures, 9 tables.

(Supplementary Information: 12 text pages, 3 figures, 7 tables)

Abstract

Susceptibility artifacts (SAs), which are inevitable for modern diffusion brain MR images with single-shot echo planar imaging (EPI) protocols in wide large-scale neuroimaging datasets, severely hamper the accurate detection of the human brain white matter structure. While several conventional and deep-learning based distortion correction methods have been proposed, the correction quality and model generality of these approaches are still limited. Here, we proposed the SACNet, a flexible SAs correction (SAC) framework for brain diffusion MR images of various phase-encoding EPI protocols based on an unsupervised learning-based registration convolutional neural network. This method could generate smooth diffeomorphic warps with optional neuroanatomy guidance to correct both geometric and intensity distortions of SAs. By employing near 2000 brain scans covering neonatal, child, adult and traveling participants, our SACNet consistently demonstrates state-of-the-art correction performance and effectively eliminates SAs-related multicenter effects compared with existing SAC methods. To facilitate the development of standard SAC tools for future neuroimaging studies, we also created easy-to-use command lines incorporating containerization techniques for quick user deployment.

Keywords: diffusion MRI, deep learning, susceptibility artifact correction, diffeomorphic registration

1. Introduction

Diffusion MRI (dMRI) provides a unique opportunity to noninvasively detect human brain white matter (WM) *in vivo* (Hagmann, 2005; Lerch et al., 2017; Sporns et al., 2005), which is highly significant for modern neuroscience and clinical brain studies. To achieve high spatial resolution and diffusion angular resolution, dMRI sequences commonly employ the echo planar imaging (EPI) technique (Biswal et al., 1995; Turner et al., 1990; Warach et al., 1995), which has a fast imaging speed and has been widely employed in various large neuroimaging projects, such as UK Biobank (Littlejohns et al., 2020), the Human Connectome Project (HCP) (Glasser et al., 2013) and the Lifespan Human Connectome Project Development (HCP-D) (Somerville et al., 2018). However, EPI, especially single-shot EPI, is substantially affected by susceptibility artifacts (SAs), resulting in severe geometric and intensity distortions (Andersson et al., 2003; Jezzard and Balaban, 1995), which largely confound accurate measurements of brain WM from the microstructure level to the whole-brain connectome level (Tax et al., 2022). Moreover, recent evidence from multicenter datasets has shown that SAs lead to the largest inconsistency in brain connectivity measurements across scan centers (Yamashita et al., 2019). Thus, developing a high-quality susceptibility artifact correction (SAC) approach is still an ongoing task for dMR brain image processing.

Many conventional methods for solving the SAC problem have been proposed. The most popular approaches generally use two frameworks: the field map method (single phase encoding, single-PE) and the inverse phase encoding (inverse-PE) based method. Both of them depend on specific EPI protocol designs. The field map approach requires an additional scan of raw magnetic field inhomogeneity (called field map) (Jezzard and Balaban, 1995; Reber et al., 1998). SAs are corrected by translating the field map into local voxel shifts. The inverse-PE approach relies on two PE-opposite EP images to capture complementary signals along inversed distortion directions (Andersson et al., 2003; Bowtell et al., 1994; Hédouin et al., 2017; Holland et al., 2010; Irfanoglu et al., 2015; Ruthotto et al., 2012). SAs are corrected by finding an ideal “middle” estimation between two inversed distorted images through iterative registration optimizations. The most recognized method of the inverse-PE approach is Topup in FSL software (Andersson et al., 2003), which presents a least-squares estimation of opposing undistorted images and shows better performance than the field map framework. However, these methods still suffer from common drawbacks, such as the narrow applicability that is restricted to specific sequence designs, limited

performance due to cumulative errors during iterative registration, and the rather long computation time.

Recently, new SAC methods utilizing convolutional neural networks (CNNs) have emerged and enabled faster and superior SAC performance than the traditional method for EPI sequences in various protocols. These methods can be mainly classified into two categories: supervised synthetic models (Hu et al., 2020; Ye et al., 2023) and unsupervised registration models (Duong et al., 2020b; Qiao and Shi, 2021; Zahneisen et al., 2020). The former models allow for SAC on single-PE images without field map. They employ additionally collected distortion-free brain images in specialized MRI protocols, such as point-spread-function (PSF)–encoded EP images, as training labels (Hu et al., 2020; Ye et al., 2023). Such supervised approaches largely depend on the feature distribution of training images (Fu et al., 2020) and thus are essentially limited when facing brain images with heterogeneous appearance, such as developmental brain scans or multicenter scans. The latter models are mainly designed for inverse-PE images (Duong et al., 2020b; Qiao and Shi, 2021; Zahneisen et al., 2020). This approach can obtain a common representation of spatial mapping between the inversed distorted brain images via the training process and thus can avoid individual iterative registrations (Balakrishnan et al., 2019). The unsupervised training process also brings high generalization ability, which is critical for robust SAC performance on various brain dMRI protocols. However, several limitations still exist for such models: 1) a compatible framework for both single- and inversed-PE type datasets is lacking, especially when facing multicenter datasets with different PE designs; 2) failure to ensure diffeomorphic transformations can lead to artificial warps during image registration; 3) prior neuroanatomical information from structural MR images is underestimated; and 4) the single-resolution strategy hampers model convergence.

To fill these gaps, we proposed SACNet, an unsupervised learning-based registration method for SA correction with the following innovations:

(i). We established a flexible mathematical correction framework for addressing the SAC problem in both inverse-PE and single-PE EPI protocols.

(ii). We proposed a diffeomorphic preservation function by modifying the Woods-Saxon potential function to restrict the generated deformation fields within a diffeomorphic solver space.

(iii). We designed an intensity-irrelevant loss function that is suitable for both T1w and T2w brain images to introduce anatomical priors for recovering cortical morphological details in severely distorted brain areas.

(iv). We devised coarse-to-fine (CTF) training and inference protocols to accelerate the learning process, leading to satisfactory model convergence.

By employing 1954 dMRI brain scans covering neonatal, child and adult populations and traveling subjects from multiple centers, we found that the proposed SACNet approach robustly outperforms both conventional and deep-learning based methods in all datasets with significantly improved correction performance, reduced multicenter effects, and low computational costs. We integrated our models into a unified pipeline and released it online at https://github.com/RicardoZiTseng/SACNet. This paper is organized as follows. In Section 2, we describe the detailed design of our approach. In Section 3, we introduce the experimental settings, including datasets, evaluation metrics and baselines. In Section 4, we present the experimental results for various datasets. In Section 5, we discuss the conclusions based on the experimental results.

2. Methods

In this section, we first present an overview of our SACNet framework (Section 2.1, Fig. 1). Then, we introduce the network architecture used in SACNet and describe the details of the differentiable EPI warp module (Section 2.2), the mathematical optimization functions (Section 2.3), the formulated optimization model and its variants for different PE protocols (Section 2.4), and the CTF training and inference approach (Section 2.5). Finally, we describe the dMRI preprocessing pipeline for SACNet (Section 2.6).

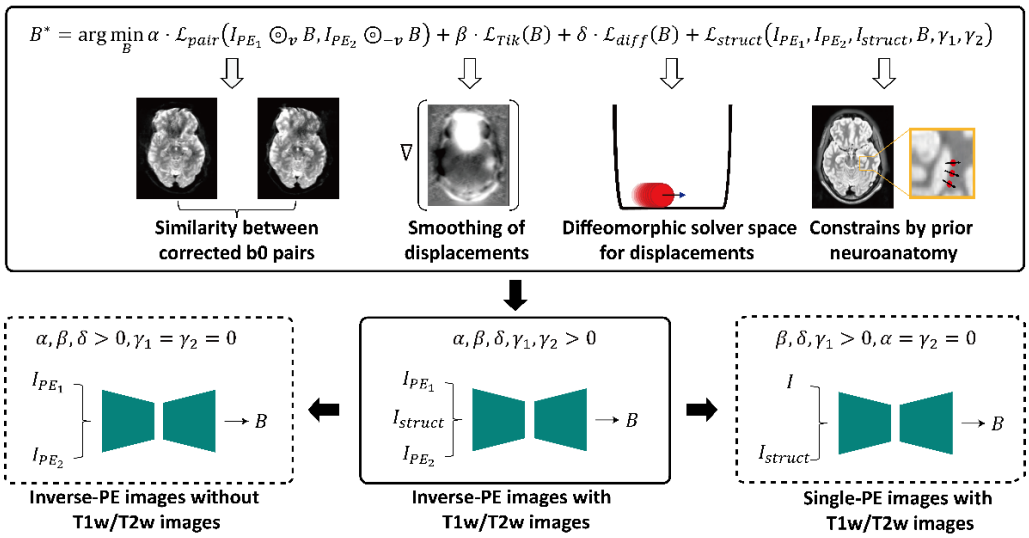
2.1. Overview

The mathematical framework and a representative flowchart of SACNet are illustrated in Fig. 1A. We designed an integrated optimization function to solve the SAC problem, including a pairwise dissimilarity loss function $\mathcal{L}_{pair}$, a Tikhonov regularization function $\mathcal{L}_{Tik}$ for estimating smooth inhomogeneity fields, a diffeomorphism preservation regularization function $\mathcal{L}_{diff}$ for guaranteeing diffeomorphic inhomogeneity fields, and a prior neuroanatomical information loss $\mathcal{L}_{struct}$ for incorporating prior neuroanatomical information. This integrated optimization function could be transformed into simple versions to make SACNet compatible with different types of PE

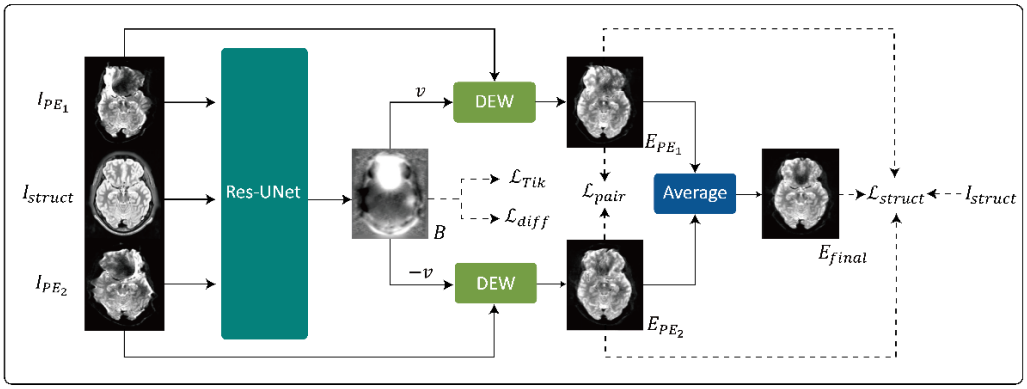
protocols by adjusting hyperparameters.

To implement this framework, we considered the approach for solving the SAC problem of an inverse-PE dataset (with T2w images as the neuroanatomical prior) as an example (Fig. 1B). Specifically, we employed Res-UNet to model the mapping from I_{PE_1}, I_{PE_2} and I_{struct} to B: $f_\theta\left(I_{PE_1}, I_{PE_2}, I_{struct}\right) = B$, in which I_{PE_1} and I_{PE_2} are the uncorrected image pair along the inverse-PE directions, I_{struct} is the structural image rigidly registered to I_{PE_1} and I_{PE_2}, B is the generated inhomogeneity field needed to remove SAs, and θ represents the network parameters. A differentiable EPI warp (DEW) module was designed to apply B to remove the SAs in I_{PE_1} and I_{PE_2} and obtain E_{PE_1} and E_{PE_2}, which are the corrected images along the two PE directions. Finally, we combined E_{PE_1} and E_{PE_2} based on the geometric average to generate the final corrected image E_{final}. For faster and better training convergence, we designed CTF training and inference protocols, as shown in Fig. 1C. The two protocols utilize several identical networks for training and inference at multiple resolution levels. Starting from the second level, the initial inhomogeneity field is upsampled based on the field calculated in the previous level, and the next fields are estimated at progressively finer levels.

A. The Mathematical Framework for SAC problem



B. The Overall Framework for Inverse-PE images with T2w images



C. The Coarse-to-fine SAC Training and Inference Protocol

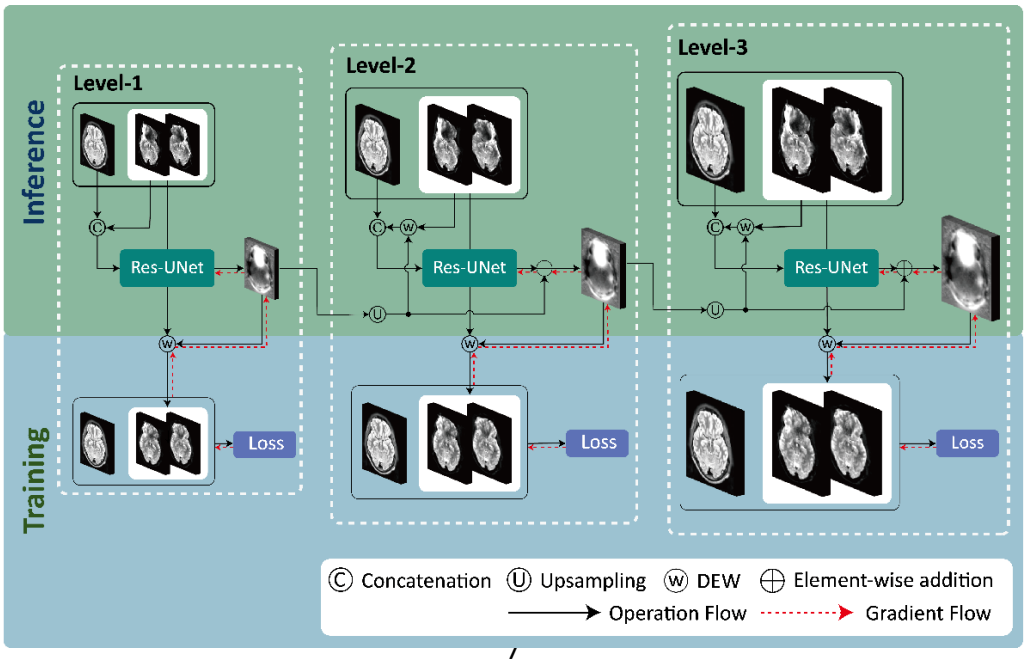


Fig. 1. The proposed mathematical framework and an implementation flowchart of SACNet. (A) The first row presents the integrated mathematical optimization function for solving the SAC problem, including a pairwise dissimilarity loss function $\mathcal{L}_{pair}$, a Tikhonov regularization function $\mathcal{L}_{Tik}$, a diffeomorphism preservation regularization function $\mathcal{L}_{diff}$, and a prior neuroanatomical information loss $\mathcal{L}_{struct}$. The second row shows that the proposed optimization model could be transformed into two simpler models to make SACNet compatible with different types of PE protocols by adjusting the hyperparameters in the optimization function. (B) The example implementation framework of SACNet with inverse-PE b0 images and T2w images as inputs is shown. All input images were sent to Res-UNet to map the inhomogeneity field B for correcting SAs. The solid line represents the data flow in the network, and the dashed line represents the participation in the loss function calculation. (C) The implementations of CTF SAC training and inference protocols for the model presented in (B). We used a series of identical networks to simulate the SAC process in the multiresolution schema. The blue part illustrates the optimization of network parameters during the training stage, and the green part illustrates the data flow during the inference stage.

2.2. Res-UNet architecture and the differentiable EPI warp module

We utilized the Res-UNet architecture to parameterize f_θ. The Res-UNet model consisted of an encoder-decoder with skip connections linking the encoder and decoder paths, and residual blocks were used to construct the whole network architecture. The implementation details of Res-UNet are described in SI-4.

Differentiable warping of raw EPI images was required to calculate the gradients during the backpropagation process. Thus, we designed a DEW module based on the spatial transformer network (Jaderberg et al., 2015) and simultaneously implemented geometry correction and intensity correction. Following the interpolator model described in previous studies (Andersson et al., 2003; Chang and Fitzpatrick, 1992; Holland et al., 2010; Studholme et al., 2000), the DEW module first resampled the given image I with inhomogeneity field B to remove geometry-related SAs and then multiplied the resampled I with the Jacobian determinant of B to remove intensity-related SAs. The specific calculation procedure is described as follows. For each image I, we first computed the voxel location $\boldsymbol{p}' = \boldsymbol{p} + B(\boldsymbol{p})\boldsymbol{v}$ for each voxel $\boldsymbol{p}$ in image I. Since the voxel intensity is defined at discrete integer locations and SAs only occur along the PE direction, we linearly interpolated the values for the left-right neighboring voxels along the PE direction $\boldsymbol{v}$:

$$I(\boldsymbol{p} + B(\boldsymbol{p})\boldsymbol{v}) = \frac{(\boldsymbol{p} + B(\boldsymbol{p}) - \boldsymbol{p}_l)\boldsymbol{v}}{(\boldsymbol{p}_r - \boldsymbol{p}_l)\boldsymbol{v}} \cdot \big(I(\boldsymbol{p}_r) - I(\boldsymbol{p}_l)\big) + I(\boldsymbol{p}_l), \qquad \forall \boldsymbol{p} \in \Omega \tag{1}$$

where $\boldsymbol{p}_r$ and $\boldsymbol{p}_l$ are the right and left neighbors of voxel $\boldsymbol{p}$ along the PE direction $\boldsymbol{v}$. Then, we

multiplied Eq. (1) by the Jacobian determinant of $\boldsymbol{B}$ to redistribute the intensity as follows:

$$E(\boldsymbol{p}) = I \circledcirc_{\boldsymbol{v}} B(\boldsymbol{p}) = I(\boldsymbol{p} + B(\boldsymbol{p})\boldsymbol{v}) \cdot clamp\left(\left(1 + \partial_{\boldsymbol{v}} B(\boldsymbol{p})\right)\right), \quad \forall \boldsymbol{p} \in \Omega \tag{2}$$

where $\left(1 + \partial_{\boldsymbol{v}} B(\boldsymbol{p})\right)$ in Eq. (2) is the Jacobian determinant of the transformation $\boldsymbol{p} \rightarrow \boldsymbol{p} + B(\boldsymbol{p})\boldsymbol{v}$ (see detailed derivation in SI-1), and $clamp(x) = max(x, 0)$ is used to prevent multiplication with a negative value.

2.3. Optimization function construction

For images with inverse-PE designs, the distorted image pair I_{PE_1} and I_{PE_2} is inversely affected by the same inhomogeneity field B along the opposite directions $\boldsymbol{v}$ and $-\boldsymbol{v}$ (Holland et al., 2010; Ruthotto et al., 2012); thus, the corrected images E_{PE_1} and E_{PE_2} were calculated as follows according to Eq. (2):

$$\begin{cases} E_{PE_1}(\boldsymbol{p}) = I_{PE_1} \circledcirc_{\boldsymbol{v}} B(\boldsymbol{p}) = I_{PE_1}(\boldsymbol{p} + B(\boldsymbol{p})\boldsymbol{v}) \cdot clamp\left(\left(1 + \partial_{\boldsymbol{v}} B(\boldsymbol{p})\right)\right) \\ E_{PE_2}(\boldsymbol{p}) = I_{PE_2} \circledcirc_{-\boldsymbol{v}} B(\boldsymbol{p}) = I_{PE_2}(\boldsymbol{p} - B(\boldsymbol{p})\boldsymbol{v}) \cdot clamp\left(\left(1 - \partial_{\boldsymbol{v}} B(\boldsymbol{p})\right)\right) \end{cases}, \quad \forall \boldsymbol{p} \in \Omega \tag{3}$$

Theoretically, we can find one solution B^* that leads to identical E_{PE_1} and E_{PE_2}; thus, the optimization problem can be formulated as:

$$\begin{aligned} B^* &= \arg\min_B \mathcal{L}_{pair}\left(I_{PE_1} \circledcirc_{\boldsymbol{v}} B, I_{PE_2} \circledcirc_{-\boldsymbol{v}} B\right) \\ &= \arg\min_B \mathcal{L}_{pair}\left(E_{PE_1}, E_{PE_2}\right) \\ &= \arg\min_B \frac{1}{|\Omega|} \sum_{\boldsymbol{p} \in \Omega} \left(E_{PE_1}(\boldsymbol{p}) - E_{PE_2}(\boldsymbol{p})\right)^2 \end{aligned} \tag{4}$$

where $\mathcal{L}_{pair}$ adjusts the pairwise dissimilarity between the estimated E_{PE_1} and E_{PE_2}. Notably, all image volumes are defined over a 3D spatial domain $\Omega \subset \mathbb{R}^3$, and $|\Omega|$ represents the number of elements in Ω.

However, previous studies have noted that seeking B^* by optimizing Eq. (4) generally leads to an ill-posed problem (Balakrishnan et al., 2019; Duong et al., 2020a; Ruthotto et al., 2012). Thus, in this paper, we introduced two regularization functions ($\mathcal{L}_{Tik}$ and $\mathcal{L}_{diff}$) and one additional loss function ($\mathcal{L}_{struct}$) to constrain B and construct the optimization function for solving the SAC problem:

$$\begin{aligned} B^* = \arg\min_B \alpha \cdot \mathcal{L}_{pair}\left(I_{PE_1} \circledcirc_{\boldsymbol{v}} B, I_{PE_2} \circledcirc_{-\boldsymbol{v}} B\right) + \beta \cdot \mathcal{L}_{Tik}(B) + \delta \cdot \mathcal{L}_{diff}(B) \\ + \mathcal{L}_{struct}\left(E_{PE_1}, E_{PE_2}, E_{final}, I_{struct}, \gamma_1, \gamma_2\right) \end{aligned} \tag{5}$$

where α, β, δ, γ_1 and γ_2 are hyperparameters used to determine the contribution of each component in Eq. (5). In addition, $\mathcal{L}_{Tik}$, $\mathcal{L}_{diff}$, and $\mathcal{L}_{struct}$ denote the Tikhonov regularization function, diffeomorphism preservation regularization function and prior neuroanatomical information loss function, respectively, which are defined in the subsequent subsections.

2.3.1. Tikhonov regularization

$\mathcal{L}_{Tik}(B)$ was used as a prior constraint on the smoothness of field B using a Tikhonov regularizer based on the spatial gradient of B:

$$\mathcal{L}_{Tik}(B) = \frac{1}{|\Omega|}\sum_{\boldsymbol{p}\in\Omega}\|\nabla B(\boldsymbol{p})\|^2 \tag{6}$$

Following the implementations in (Balakrishnan et al., 2019), for $\nabla B(\boldsymbol{p}) = \left(\frac{\partial B(p_x)}{\partial x}, \frac{\partial B(p_y)}{\partial y}, \frac{\partial B(p_z)}{\partial z}\right)$, we approximated $\frac{\partial B(p_x)}{\partial x} \approx B\left(\left(p_x+1, p_y, p_z\right)\right) - B\left(\left(p_x, p_y, p_z\right)\right)$, and we used similar approximations for $\frac{\partial B(p_y)}{\partial y}$ and $\frac{\partial B(p_z)}{\partial z}$.

2.3.2. Diffeomorphism preservation regularization

To guarantee the diffeomorphism property of the inhomogeneity field, we proposed a diffeomorphism preservation regularization function by modifying a potential well function. Specifically, in terms of intensity, we expected that the signals of the voxels at the same position in E_1 and E_2 would both be positive, which requires the following:

$$\begin{cases} 1 + \partial_{\boldsymbol{v}} B(\boldsymbol{p}) > 0 \\ 1 - \partial_{\boldsymbol{v}} B(\boldsymbol{p}) > 0 \end{cases}, \forall \boldsymbol{p} \in \Omega \tag{7}$$

This is equivalent to:

$$-1 < \partial_{\boldsymbol{v}} B(\boldsymbol{p}) < 1, \forall \boldsymbol{p} \in \Omega \tag{8}$$

In terms of the geometry, we expected the relative positions of adjacent voxels to remain the same before and after resampling, which guarantees no folding areas during the transformation, as shown in Fig. 2A. For example, along the direction $\boldsymbol{v} = (1,0,0)$, the displacements of point $\boldsymbol{p}_0 = (\boldsymbol{p_x}, \boldsymbol{p_y}, \boldsymbol{p_z})$ and its neighbor $\boldsymbol{p}_{+1} = (\boldsymbol{p_x}+1, \boldsymbol{p_y}, \boldsymbol{p_z})$ are $B(\boldsymbol{p}_0)\boldsymbol{v}$ and $B(\boldsymbol{p}_{+1})\boldsymbol{v}$, respectively. These displacements are similar along the direction $-\boldsymbol{v} = (-1,0,0)$. To prevent the folding of space at point x, the new spatial positions should follow:

$$\begin{cases} \boldsymbol{p}_0 + B(\boldsymbol{p}_0)\boldsymbol{v} < \boldsymbol{p}_{+1} + B(\boldsymbol{p}_{+1})\boldsymbol{v} \\ \boldsymbol{p}_{-1} - B(\boldsymbol{p}_{-1})\boldsymbol{v} < \boldsymbol{p}_0 - B(\boldsymbol{p}_0)\boldsymbol{v} \end{cases}, \ \forall \boldsymbol{p} \in \Omega \tag{9}$$

We can also obtain Eq. (8) from Eq. (9).

To ensure that the generated field B satisfies the constraint function shown in Eq. (8), we expected that when $\partial_v B(p)$ approached -1 or 1, the loss function increased substantially, and when $\partial_v B(p)$ remained between -1 and 1, the loss function remained small. To this end, we designed the diffeomorphism preservation function (DPF) as follows:

$$\mathcal{L}_{diff}(B) = \frac{1}{|\Omega|} \sum_{\boldsymbol{p}} \phi\big(\partial_v B(\boldsymbol{p})\big) * \big(\partial_v B(\boldsymbol{p})\big)^2 \tag{10}$$

where $\phi(\cdot)$ is the potential well function modified from the Woods-Saxon potential function widely used in nuclear physics (Erkol and Demiralp, 2007):

$$\phi(x) = \left(1 - \frac{1}{1 + \exp[(|x| - 1)/\sigma]}\right) \tag{11}$$

where σ is a customized parameter. Fig. 2B shows the curves of $\phi(x)$ (the left subgraph) in terms of $\partial_{\mathbf{v}} B(\boldsymbol{p})$ and the derivative (the right subgraph) with respect to $\partial_{\mathbf{v}} B(\boldsymbol{p})$. The figure shows that the value of $\phi(x)$ increases substantially as $|\partial_v B(\boldsymbol{p})| \to 1$, which suggests that $\phi(x)$ can sensitively suppress the voxels that do not obey the constraint defined in Eq. (8), thereby constraining the inhomogeneity field B to a diffeomorphic space. Notably, $\big(\partial_v B(\boldsymbol{p})\big)^2$ is multiplied by $\phi\big(\partial_v B(\boldsymbol{p})\big)$ to prevent the gradient from vanishing when $\partial_v B(\boldsymbol{p})$ is larger than 1 or smaller than -1. We present the proof of the existence of a diffeomorphic inhomogeneity field calculated by SACNet in SI-2.

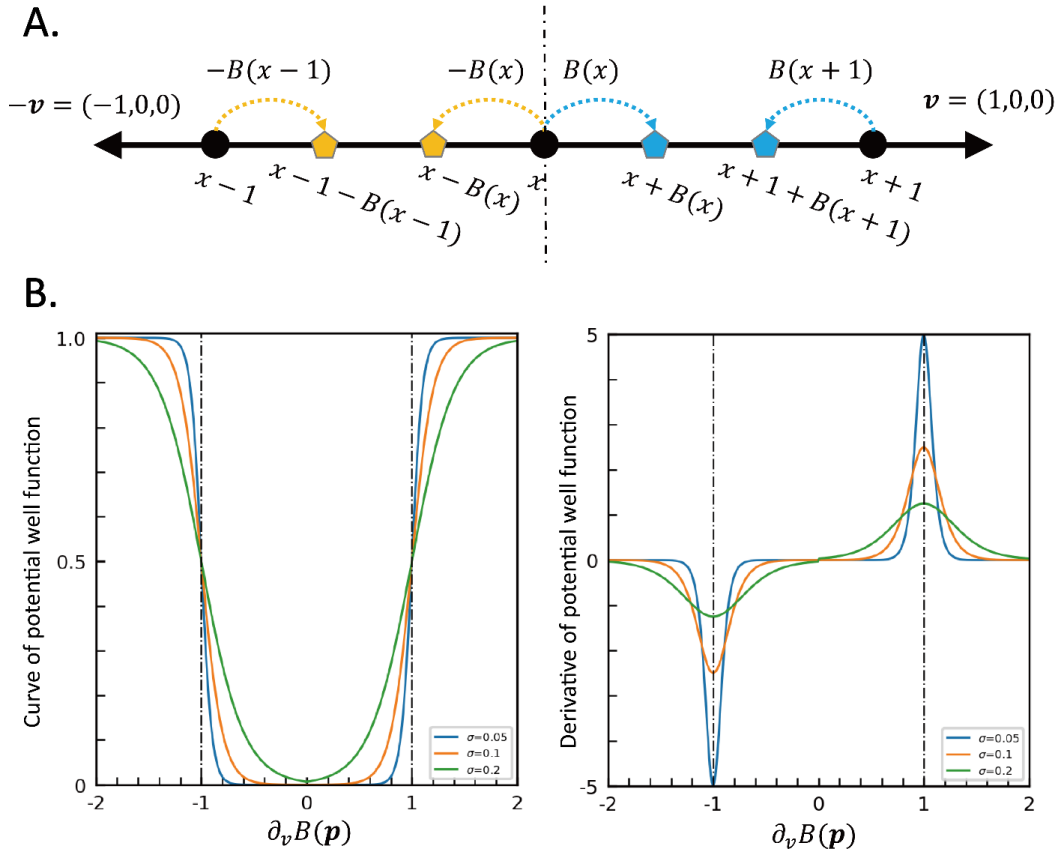


Fig. 2. Mathematical framework of the diffeomorphism preservation regularization. A) Illustration of spatial folding at location x along opposite directions $\boldsymbol{v}$ and $-\boldsymbol{v}$. Here, $x+1$ and $x-1$ denote the nearest neighbors of x, and we assume that $\boldsymbol{v} = (1,0,0)$ and B is an inhomogeneity field. (B) The left part shows the function value of $\phi(x)$ with different hyperparameters σ and the value in terms of $\partial_v B(\boldsymbol{p})$. The right part shows the derivative of $\phi(x)$ with respect to $\partial_v B(\boldsymbol{p})$.

2.3.3. Prior neuroanatomical information loss

Image noise caused by SAs hinders strict alignment of the b0 image pair, resulting in inaccurate estimation of the inhomogeneity field in severely distorted areas. To address this issue, we proposed a prior neuroanatomical information loss $\mathcal{L}_{struct}$ to incorporate accurate prior neuroanatomical information. This approach has two main benefits. First, this loss regularizes the inhomogeneity field while preserving intricate neuroanatomical morphological details. Second, it provides an additional registration target when the b0 image pair is not available.

$\mathcal{L}_{struct}$ consists of two parts: the overall shape structural similarity loss $\mathcal{L}_{str-overall}$ and the pairwise structural similarity loss $\mathcal{L}_{str-pair}$. Conceptually, $\mathcal{L}_{str-overall}$ ensures that the final corrected image E_{final} is similar to the provided structural image I_{struct}, while $\mathcal{L}_{str-pair}$ ensures

that the corrected images along each PE direction E_{PE_1} and E_{PE_2} are similar to I_{struct}. Specifically, the proposed neuroanatomy prior loss is formulated as:

$$\mathcal{L}_{struct}\left(I_{PE_1}, I_{PE_2}, I_{struct}, B, \gamma_1, \gamma_2\right) = \gamma_1 \cdot \mathcal{L}_{str-overall} + \gamma_2 \cdot \mathcal{L}_{str-pair} \tag{12}$$

with

$$\mathcal{L}_{str-overall} = \mathcal{L}_{sim}\left(E_{final}, I_{struct}\right) \tag{13}$$

and

$$\mathcal{L}_{str-pair} = \mathcal{L}_{sim}\left(E_{PE_1}, I_{struct}\right) + \mathcal{L}_{sim}\left(E_{PE_2}, I_{struct}\right) \tag{14}$$

where γ_1 and γ_2 are two user-defined hyperparameters, and $\mathcal{L}_{sim}$ in Eq. (13) and Eq. (14) represents a similarity metric.

We anticipated that SACNet would not be limited to the MR modality of structural inputs. Therefore, instead of relying on absolute intensity-relevant similarity metrics, such as the mean square error (MSE) and local cross-correlation (LCC), we employed a gradient-based similarity metric, namely, the normalized gradient field (NGF), as $\mathcal{L}_{sim}$. The NGF determines the geometric resemblance between any points in an image by computing local gradients; thus, this metric is independent of the absolute image intensity (Haber and Modersitzki, 2007). Let $\nabla X_{\boldsymbol{p}}$ be the intensity change gradient at point $\boldsymbol{p} \in \Omega$ in image X and ϵ be a user-defined parameter that prevents divide-by-zero errors. Then, the NGF measure at any point $\boldsymbol{p}$ in image X can be defined as:

$$\widetilde{\nabla} X(\boldsymbol{p}) = \frac{\nabla X(\boldsymbol{p})}{\sqrt{\|\nabla X(\boldsymbol{p})\|^2 + \epsilon^2}} \tag{15}$$

The difference between two images X and Y can be measured by calculating the angles between the NGF vectors at all points in the image domain, which can be formulated as follows:

$$\mathcal{L}_{sim}(X, Y) = \frac{1}{|\Omega|} \sum_{\boldsymbol{p} \in \Omega} \left(1 - \langle \widetilde{\nabla} X(\boldsymbol{p}), \widetilde{\nabla} Y(\boldsymbol{p}) \rangle\right)^2 \tag{16}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner dot-product operation. The value of $\mathcal{L}_{\mathrm{NGF}}(X, Y)$ is positive, and the smaller the value of $\mathcal{L}_{\mathrm{sim}}(X, Y)$ is, the more similar the two images are.

2.4. The formulated optimization model and its variants

To handle the different imaging protocols in various existing neuroimaging datasets, the optimization model formulated in Eq. (5) can be transformed into two different forms, as shown

in Fig. 1A: a) When no structural images are available (image set $\{I_{PE_1}, I_{PE_2}\}$), the model can be transformed to use Eq. (17) by setting γ_1 and γ_2 to 0, as illustrated in the first column in the second row of Fig. 1A:

$$B^* = \arg\min_B \left[\alpha \cdot \mathcal{L}_{pair}\left(E_{PE_1}, E_{PE_2}\right) + \beta \cdot \mathcal{L}_{smo}(B) + \delta \cdot \mathcal{L}_{diff}(B)\right] \tag{17}$$

b) When only single-PE images are available (only one single-direction distorted image and one structural image, with the image set $\{I, I_{struct}\}$), the model can be transformed to use Eq. (18) by setting α and γ_2 to 0, as illustrated in the third column in the second row of Fig. 1A:

$$B^* = \arg\min_B \left[\beta \cdot \mathcal{L}_{smo}(B) + \delta \cdot \mathcal{L}_{diff}(B) + \gamma_1 \cdot \mathcal{L}_{sim}\left(E_{final}, I_{struct}\right)\right] \tag{18}$$

In this situation, $E_{final} = I \circledcirc_{\boldsymbol{v}} B$ denotes the image corrected based on the distorted image I along the single-PE direction $\boldsymbol{v}$. In addition, the potential well function in Eq. (11) can be reformulated as:

$$\phi(x) = \left(1 - \frac{1}{1 + \exp[(-1 - x)/\sigma]}\right) \tag{19}$$

To use a neural network to predict the inhomogeneity field, the overall loss varies for the three types of image sets $\{I_{PE_1}, I_{PE_2}, I_{struct}\}$, $\{I_{PE_1}, I_{PE_2}\}$ and $\{I, I_{struct}\}$ for one subject as follows:

$$\begin{aligned}\mathcal{L}_{tot_1}\left(I_{PE_1}, I_{PE_2}, I_{struct}, B\right) = \alpha \cdot \mathcal{L}_{pair}\left(E_{PE_1}, E_{PE_2}\right) + \beta \cdot \mathcal{L}_{smo}(B) \\ +\delta \cdot \mathcal{L}_{diff}(B) + \mathcal{L}_{struct}\left(I_{PE_1}, I_{PE_2}, I_{struct}, B, \gamma_1, \gamma_2\right)\end{aligned} \tag{20}$$

$$\mathcal{L}_{tot_2}(I_{PE_1}, I_{PE_2}, B) = \alpha \cdot \mathcal{L}_{pair}\left(E_{PE_1}, E_{PE_2}\right) + \beta \cdot \mathcal{L}_{smo}(B) + \delta \cdot \mathcal{L}_{diff}(B) \tag{21}$$

$$\mathcal{L}_{tot_3}(I, I_{struct}, B) = \beta \cdot \mathcal{L}_{smo}(B) + \delta \cdot \mathcal{L}_{diff}(B) + \gamma_1 \cdot \mathcal{L}_{sim}\left(E_{final}, I_{struct}\right) \tag{22}$$

2.5. Coarse-to-fine (CTF) SAC training and inference protocols

To improve the training process and prevent falling into local minima, we designed CTF training and inference protocols for SACNet, as illustrated in Fig. 1C. The CTF training protocol aims to train multiple networks at N_s scale levels, with each model estimating the residual inhomogeneity field at the corresponding scale. Specifically, we first trained the network at the coarsest scale level and then progressively trained the networks at the each subsequent scale level to solve the SAC problem at finer scale levels. This training procedure was repeated until the model was trained at the finest level. The CTF inference protocol aimed to generate the estimated inhomogeneity field based on the training protocol using multiple trained networks. At each scale level s, we downsampled the image set by $2^{N_s - s}$ times and upsampled the inhomogeneity field $B^{(s-1)}$ 2 times.

Then, we fed the downsampled image set into the network at the current level to obtain the residual inhomogeneity field $\Delta B^{(s)}$. The inhomogeneity field at the current level was calculated by summing the upsampled field and the residual field. The pseudocodes for the training and inference protocols are presented in Algorithms 1 and 2, respectively.

Algorithm 1. Coarse-to-fine SAC training protocol of SACNet, as depicted in Fig. 1C.

Input: Training datasets $\{D_t^i\}_{i=1}^{N_t}$, validation dataset $\{D_v^i\}_{i=1}^{N_v}$. For each subject's image set, $D^i = \{I_{PE_1}^i, I_{PE_2}^i, I_{struct}^i\}$.

N_B: batch size, N_S: the number of scale levels. $\alpha^{(s)}, \beta^{(s)}, \gamma_1^{(s)}, \gamma_2^{(s)}, \delta^{(s)}$: the hyperparameters in Eq. (20) for scale level s. $E^{(s)}$: the number of epochs for the network with scale level s.

Initialization: Initialize N_S network set $F_{1:N_S} = \{f_{\theta_1}, f_{\theta_2}, \dots, f_{\theta_{N_S}}\}$ for N_S scales. Set $B^{(0)} = \mathbf{0}$.

for $s_{curr} \leftarrow 1$ **to** N_S **do**
  Set $metric_{min} = inf$ and initialize $\theta^{(s_{curr})}$ as θ_{best} if s is not equal to 1;
  for $\boldsymbol{e} \leftarrow 1$ to $E^{(s_{curr})}$ do
    repeat
      Randomly select batches of training image sets $\{D_t^i\}$ of size N_B;
      Downsample image set $\{D_t^i\}$ $2^{N_s - s_{curr}}$ times for $i = 1, \dots, N_B$;
      Compute the inhomogeneity field $\{B^{i(s_{curr}-1)} = SACNet_infer(D_t^i, \{f_{\theta_1}, \dots, f_{\theta_{s_{curr}-1}}\}, N_s)\}$ of the previous level for $i = 1, \dots, N_B$;
      Upsample $\{B^{i(s_{curr}-1)}\}$ 2 times for $i = 1, \dots, N_B$;
      Compute the estimated residual inhomogeneity field $\{\Delta B^{i(s_{curr})}\}$ by inputting $\{D_t^i\}$ into the current level's network $f_{\theta^{(s_{curr})}}$ $i = 1, \dots, N_B$;
      Compute the estimated inhomogeneity field of the current level $\{B^{i(s_{curr})}\}$ by adding $\{B^{i(s_{curr}-1)}\}$ and $\{\Delta B^{i(s_{curr})}\}$ for $i = 1, \dots, N_B$;
      Compute the loss $\mathcal{L}_{tot}^{(s_{curr})}$ according to Eq. (20) and use the Adam optimizer to update the model parameters $\theta^{(s_{curr})}$;
    until all training data have been selected
    Compute the inhomogeneity field $\{(B^i, E_{PE_1}^i, E_{PE_2}^i) = SACNet_infer\left(D_v^i, \{f_{\theta_1}, \dots, f_{\theta_{s_{curr}}}\}, N_s\right)\}$ for $i = 1, \dots, N_B$;
    Compute the sum of the mean square error $metric^{(e)}$ between $E_{PE_1}^i$ and $E_{PE_2}^i$ for $i = 1, \dots, N_B$;
    Set $metric_{min} \leftarrow metric^{(e)}$ and save model parameters $\theta^{(s_{curr})}$ to θ_{best} **if** $metric^{(e)}$ is smaller than $metric_{min}$;
    Unfreeze all previous levels' model parameters if the model has been trained for $E^{(s_{curr})}/3$ epochs at each scale level.
  end for
  Freeze the current model parameters $\theta^{(s_{curr})}$.
end for

Return: trained network set $F_{1:N_S} = \{f_{\theta_1}, f_{\theta_2}, \dots, f_{\theta_{N_S}}\}$.

Algorithm 2. Coarse-to-fine SAC inference protocol of SACNet, as depicted in Fig. 1C.

Input: image set $D = \{I_{PE_1}, I_{PE_2}, I_{struct}\}$; trained model set $F_{1:s} = \{f_{\theta_1}, f_{\theta_2}, \dots, f_{\theta_s}\}$; and total number of scale levels N_S.

Initialization: set $B^{(0)} = \mathbf{0}$.

function $SACNet_infer(D, F_{1:S}, N_S)$

$N_{models} = length(F_{1:S})$;

for $s_{curr} \leftarrow 1$ **to** N_{models} **do**

Downsample image set $\{I_{PE_1}, I_{PE_2}, I_{struct}\}$ $2^{N_S - s_{curr}}$ times;

Upsample previous level's estimated inhomogeneity field $B^{(s_{curr}-1)}$ 2 times;

Compute the estimated residual inhomogeneity field of the current level $\Delta B^{(s_{curr})} = f_{\theta_{s_{curr}}}(I_{PE_1}^{(s_{curr})} \odot_{\boldsymbol{v}} B^{(s_{curr}-1)}, I_{PE_2}^{(s_{curr})} \odot_{-\boldsymbol{v}} B^{(s_{curr}-1)}, I_{struct}^{(s_{curr})})$;

Compute the estimated inhomogeneity field of the current level $B^{(s_{curr})} = B^{(s_{curr}-1)} + \Delta B^{(s_{curr})}$;

end for

Compute the corrected images $E_{PE_1} = I_{PE_1} \odot_{\boldsymbol{v}} B^{(N_{models})}$ and $E_{PE_2} = I_{PE_2} \odot_{-\boldsymbol{v}} B^{(N_{models})}$.

return $B^{(N_{models})}$, E_{PE_1}, E_{PE_2}.

end function

Return: inhomogeneity field $B^{(N_{models})}$ and corrected image pair E_{PE_1} and E_{PE_2}.

2.6. The whole dMRI preprocessing pipeline

We introduced a practical dMRI preprocessing pipeline by integrating SACNet with the Eddy tool in FSL, which is available in our online code. The pipeline started by correcting for motion and eddy current distortions in the dMRI volumes along each PE direction using the FSL Eddy tool. Next, the structural (T1w and T2w) images, as well as all negative and positive PE scans, were rigidly coregistered using the Flirt tool in FSL, with the first b0 image serving as the target. Then, the aligned positive and negative b0 images were input into the trained model to estimate the inhomogeneity field, which is subsequently used to remove SAs in all diffusion weighted images (DWIs).

3. Experimental settings

3.1. Datasets

To comprehensively evaluate the SAC performance of our proposed approach, we considered multiple existing large neuroimaging datasets that contain structural and dMRI scans across age groups and acquisition protocols and assessed whether our approach could achieve state-of-the-art performance, including 1) excellent performance based on adult brain images by randomly selecting 380 adult inverse-PE dMR images from the HCP dataset (Glasser et al., 2013); 2) excellent performance based on developmental brain images by employing 444 neonatal scans from the Developing Human Connectome Project (dHCP) dataset (Makropoulos et al., 2018) and 1100 children and adolescent scans from the HCP-D dataset (Somerville et al., 2018) and the Children School Functions and Brain Development Project in China (CBD) dataset (HCP-D: 644 scans, CBD: 456 scans); and 3) low multicenter effects with excellent SAC performance by

adopting 30 scans of three healthy traveling subjects acquired at 10 scan sites from a multicenter public dataset (Multicenter) (Tong et al., 2020). Notably, the CBD dataset was also used to examine the SAC capability of SACNet based on single-PE data. The details of each dataset are listed in Table 1, and the detailed preprocessing methods for each dataset are described in SI-3.

3.2. Implementation, evaluation, and statistical methods

The proposed method was implemented in Python using the PyTorch software library (Paszke et al., 2019). Our model was trained and tested on a Linux workstation equipped with an Intel Xeon Gold 6258R CPU and a 48 GB GTX Quadro RTX 8000 GPU. We employed the Adam optimizer (Kingma and Ba, 2014) with a learning rate of 1e-4 for optimization. The specific training and inference configurations for each dataset are detailed in SI-3.

To quantitatively assess SAC performance for diffusion model fitting, we calculated several metrics based on fractional anisotropy (FA). We did not employ b0 images for the estimation because b0 images are unable to reflect correction quality in image volumes of diffusion weighting directions. These metrics included the FA-based mean squared difference (FA-MSD) between different PE directions, FA-based standard deviation (FA-STD) across multiple PE directions and FA-based structural similarity (FA-SS, local cross-correlation between FA and structural images (window size = 3)). To evaluate the diffeomorphism of the estimated inhomogeneity field, we calculated the number of folding voxels (NFV). For the HCP, HCP-D and multicenter datasets, we used the FA-MSD, FA-SS and NFV as evaluation indices. Since the dHCP dataset has four unique PE directions and does not contain a suitable number of structural images, we chose the FA-STD instead of the FA-SS as an evaluation index. For the CBD dataset, which has only one PE direction, we chose only the FA-SS and NFV as quantitative evaluation indices. The detailed calculation of these metrics is described in SI-5.

We used the paired t test on these quantitative metrics for statistical comparisons between different methods. The detailed results of the statistical analysis are summarized in SI-7.

3.3. Baselines

We adopted conventional and deep-learning based approaches that have demonstrated excellent performance as baseline models. For inverse-PE datasets, we compared SACNet with the FSL Topup (Andersson et al., 2003) and S-Net (Duong et al., 2020b) approaches. For the single-PE datasets, we compared SACNet with the Fieldmap method included in FSL software and a widely

adopted deep-learning image registration baseline, namely, VoxelMorph (Balakrishnan et al., 2019). Notably, we constrained the deformation field along the PE direction in VoxelMorph.

To evaluate the effectiveness of prior neuroanatomical information, we considered three variants of SACNet by inputting only paired b0 images (SACNet(wos)) or inputting structural images with or without paired b0 images (SACNet(T1w) for T1w image input and SACNet(T2w) for T2w image input). All three variants were evaluated based on the HCP and HCP-D datasets. For the dHCP dataset, we only trained SACNet(wos) and SACNet(T2w), as T1w images were not acquired for some neonates. For the CBD datasets, we only trained SACNet(T1w) and SACNet(T2w) without paired b0 image inputs since these datasets do not include paired b0 images. For the multicenter dataset, we only trained SACNet(wos) and SACNet(T1w), as this dataset includes only T1w structural images.

Dataset	HCP	HCP-D	dHCP	CBD	Multicenter
Number of Scans	380	644	444	456	30
Manufacturer	Siemens	Siemens	Philips	Siemens	Siemens
Platform	Customized Skyra	Prisma	Achieva	Prisma	Prisma
Magnetic Field Strength [Tesla]	3.0	3.0	3.0	3.0	3.0
Diffusion Weighted Image					
Phase Encoding Direction	RL, LR	AP, PA	RL, LR, AP, PA	PA	AP, PA
Echo Time [ms]	89.5	89.2	90	64	71
Repetition Time [ms]	5520	3230	3800	7500	5400
Image Dimension	145×174×145	140×140×92	128×128×64	112×112×70	146×146×92
Resolution [mm^3]	1.25×1.25×1.25	1.5×1.5×1.5	1.17×1.17×1.5	2×2×2	1.5×1.5×1.5
Structural Image					
Echo Time [ms]	2.14 (T1w)/565 (T2w)	1.8/3.6/5.4/7.2 (T1w)/564 (T2w)	156 (T2w)	2.98 (T1w)/564 (T2w)	2.9 (T1w)
Repetition Time [ms]	2400 (T1w)/3200 (T2w)	2500 (T1w)/3200 (T2w)	1200 (T2w)	2530 (T1w)/3200 (T2w)	5000 (T1w)
Image Dimension	260×311×260 (T1w, T2w)	208×300×320 (T1w, T2w)	290×290×203 (T2w)	256×224×192 (T1w)/320×320×256	176×240×256 (T1w)

Resolution [mm^3]	0.7×0.7×0.7 (T1w, T2w)	0.8×0.8×0.8 (T1w, T2w)	0.5×0.5×0.5 (T2w)	(T2w) 1×1×1 (T1w)/0.7×0.7×0.7 (T2w)	1.2×1×1 (T1w)

Table 1. The acquisition parameter details of each dataset. RL, LR, AP and PA denote the right-left, left-right, anterior-posterior and posterior-anterior phase-encoding directions, respectively.

4. Results

4.1. Performance on the inverse-PE adult dataset

We first used brain scans with inverse-PE protocols in the HCP dataset (300 of 380 subjects were used for training, 40 for validation and 40 for testing) to test our model. We visualized the corrected b0 images and corresponding FA maps (first and second row, Fig. 3A) and calculated the FA-based structural similarity map across the whole brain (third row, Fig. 3A) of different SAC approaches. Our SACNet models demonstrated better correction quality, particularly in the frontal cortex, than Topup and S-Net. Further quantitative comparisons showed that the SACNet(T1w) and SACNet(T2w) models exhibited significantly better performance in terms of both the FA-MSD (all $t \leq -3.678$ and all $p<0.001$, paired t test) and FA-SS (all $t \geq 3.716$ and all $p<0.001$, paired t test) metrics compared to the Topup and S-Net methods, with remarkable improvements of up to 15.9% and 5.9%, respectively (Fig. 3B and Table 2). Furthermore, compared with the SACNet(wos) model without structural images, the SACNet(T1w) and SACNet(T2w) models achieved significantly lower FA-MSD (all $t \leq -19.455$ and all $p<0.001$, paired t test) and higher FA-SS (all $t \geq 15.775$ and all $p<0.001$, paired t test) values, indicating the necessity of introducing the prior neuroanatomical information loss component. Finally, our approach showed substantial reductions in the NFV metric (from 1533.8 for Topup and 756.9 for S-Net to only 15.2 for SACNet(T2w), Table 2). More detailed statistical comparisons are shown in Supplementary Table S1.

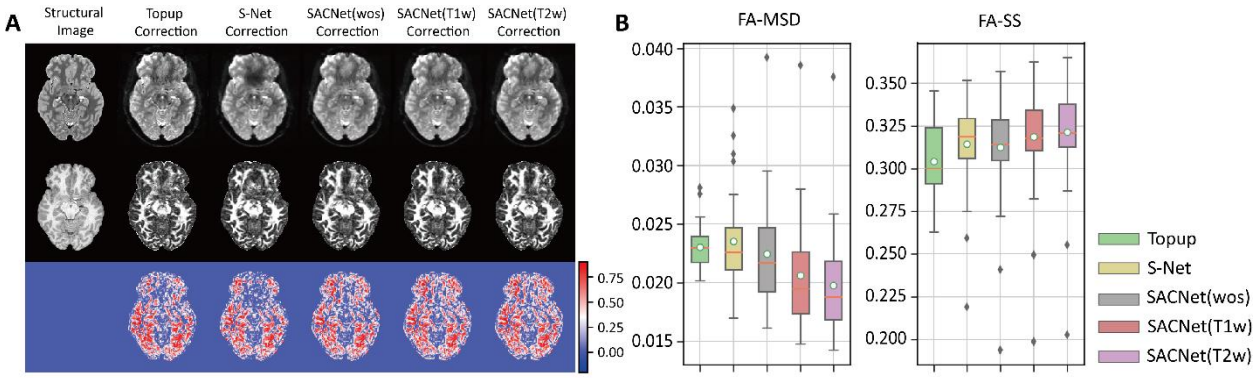


Fig. 3. SAC performance based on the HCP dataset (A) The first and second rows of the first column present

the T2w and T1w images, respectively. The other columns of the first, second and third rows present the corrected b0 images, the corrected FA maps and the FA-SS maps (local cross-correlation with window size=3 between the FA map and the T1w image) for different correction methods, respectively. Warmer colors indicate higher FA-SS values. (B) Boxplots of the FA-MSD and FA-SS values for each method. White circles indicate mean values, and coral horizontal lines indicate median values.

Method	Metrics		
	FA-MSD (1e-2)	FA-SS	NFV
Topup	2.304±0.180	0.304±0.022	1533.82±456.20
S-Net	2.353±0.377	0.314±0.026	756.98±374.07
SACNet(wos)	2.244±0.444	0.313±0.031	**13.62±27.60**
SACNet(T1w)	2.062±0.442	0.319±0.031	14.28±35.59
SACNet(T2w)	**1.978±0.420**	**0.322±0.030**	15.22±38.17

Table 2. The quantitative results based on the HCP dataset. Bold font denotes that the method performs the best among the considered methods.

4.2. Performance on the inverse-PE developmental datasets

We further conducted experiments based on children's brain scans. We randomly selected fMRI and structural MR brain images from children in the HCP-D dataset (544 of 644 subjects were used for training, 50 for validation and 50 for testing) and neonates in the dHCP dataset (364 of 444 subjects were used for training, 40 for validation and 40 for testing) to estimate the performance of SACNet.

For the HCP-D dataset, we observed that our approach obtained better correction quality at cortical boundaries in the frontal gyrus than the Topup and S-Net methods (Fig. 4A). Further quantitative analyses (Fig. 4B and Table 3) showed that SACNet(T1w) and SACNet(T2w) significantly outperformed the Topup and S-Net methods in terms of the FA-MSD (all $t \leq -15.174$ and all $p<0.001$, paired t test) and FA-SS (all $t \geq 7.483$ and all $p<0.001$, paired t test) metrics. Furthermore, our approach greatly reduced the NFV metric (Table 3). Detailed statistical results are presented in Supplementary Table S2.

For the dHCP dataset (Fig. 4C and 4D, Table 4), our SACNet(T2w) approach significantly outperformed the Topup and S-Net methods in terms of both the FA-MSD (all $t \leq -3.998$ and all $p<0.001$, paired t test) and FA-STD (all $t \leq -4.098$ and all $p<0.001$, paired t test) metrics. Detailed statistical results are presented in Supplementary Table S3.

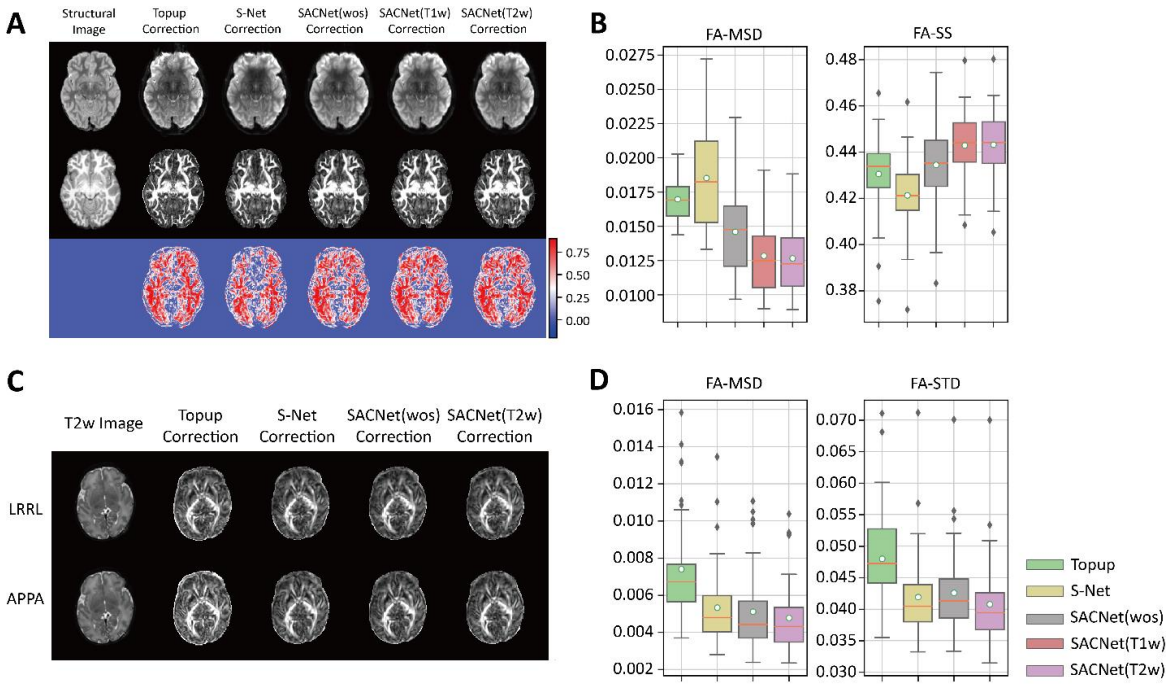


Fig. 4. (A) The first and second rows of the first column present the T2w and T1w images, respectively. The other columns of the first, second and third rows present the corrected b0 images, the corrected FA maps and the FA-SS maps (local cross-correlation with window size=3 between the FA map and the T1w image) for different correction methods, respectively. Warmer colors indicate higher FA-SS values. (B) Boxplots of the FA-MSD and FA-SS values for each method based on the HCP-D dataset. (C) The first column presents T2w images. The other columns of the first and second rows present the corrected FA maps along the LR-RL and AP-PA directions for different correction methods, respectively. (D) Boxplots of the FA-MSD and FA-STD values for each method based on the dHCP dataset. For the boxplots in (B) and (D), the white circles indicate the mean values, and the coral horizontal lines indicate the median values.

Method	Metrics		
	FA-MSD (1e-2)	FA-SS	NFV
Topup	1.698±0.151	0.431±0.015	1615.00±536.43
S-Net	1.852±0.369	0.421±0.015	1046.68±562.00
SACNet(wos)	1.459±0.314	0.434±0.016	**1.44±10.08**
SACNet(T1w)	1.285±0.252	0.443±0.014	7.28±18.56
SACNet(T2w)	**1.265±0.248**	**0.443±0.014**	6.74±22.89

Table 3. The quantitative results based on the HCP-D dataset. Bold font denotes that the method performed the best among the considered methods.

Method	Metrics			
	FA-MSD (1e-2)	FA-STD	NFV (LR-RL)	NFV (AP-PA)

Topup	0.729±0.262	0.047±0.007	1071.46±1114.40	44.80±96.46
S-Net	0.522±0.208	0.042±0.007	99.32±74.93	176.24±89.54
SACNet(wos)	0.496±0.197	0.042±0.007	3.50±11.62	0.27±1.70
SACNet(T2w)	**0.464±0.180**	**0.041±0.007**	**3.41±8.48**	**0.10±0.62**

Table 4. The quantitative results based on the dHCP dataset. Bold font denotes that the method performed the best among the considered methods.

4.3. Performance on the single-PE developmental dataset

Next, we employed brain scans from a subset of the CBD project acquired from Peking University (CBDP) (242 of 322 subjects were used for training, 40 for validation and 40 for testing) to evaluate the performance of SACNet on single-PE developmental brain images. The visual examination showed that our approach exhibited better correction quality in the frontal cortex than the Fieldmap and VoxelMorph methods (Fig. 5A). Quantitatively, SACNet(T1w) and SACNet(T2w) obtained significantly higher FA-SS values than FieldMap and VoxelMorph (all $t \geq 5.581$ and all $p<0.001$, paired t test).

To further evaluate the generalization performance of SACNet on single-PE images across centers, we used another subset acquired from Beijing Huilongguan Hospital (CBDH) (all 134 subjects were used for testing the model trained with the CBDP subset). The quantitative results were similar to what we observed with the CBDP subset (all $t \geq 9.698$ and all $p<0.001$, paired t test), indicating the robust generalization capability of SACNet with the single-PE developmental dataset.

We did not show the NFV value of each method for the two subsets because all of them were close to 0. Quantitative results are provided in Fig. 5B and Table 5. The detailed statistical results are provided in Supplementary Table S4.

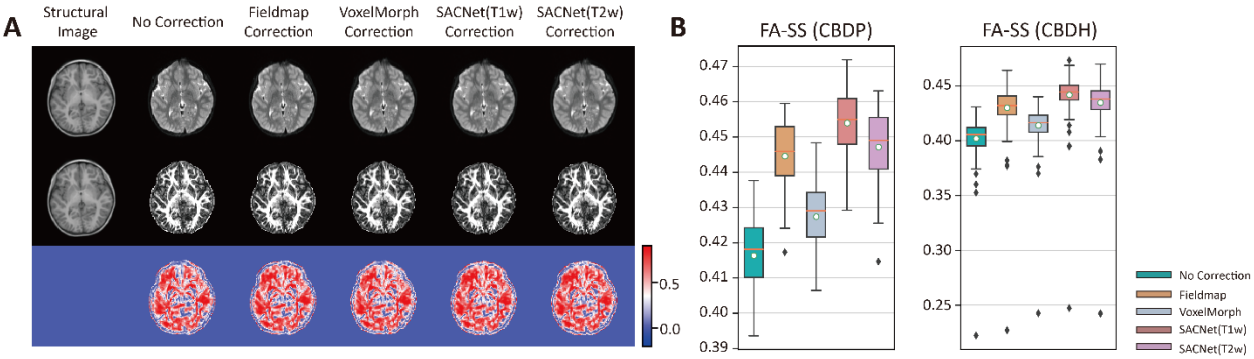


Fig. 5. (A) The visualization results based on one of the subjects in the CBDP dataset. The first column

presents the T1w image. The other columns of the first, second and third rows present the corrected b0 images, the corrected FA maps and the FA-SS maps (local cross-correlation with window size=3 between the FA map and the T1w image) for different correction methods, respectively. Warmer colors indicate higher FA-SS values. (B) Boxplots of the FA-SS values for each method based on the CBDP and CBDH subsets. The white circles indicate the mean value, and the coral horizontal lines indicate the median value.

Method	Datasets & Metrics (FA-SS)	
	CBDP	CBDH
No Correction	0.416±0.012	0.402±0.021
Fieldmap	0.445±0.011	0.430±0.023
VoxelMorph	0.427±0.010	0.414±0.020
SACNet(T1w)	**0.454±0.010**	**0.442±0.021**
SACNet(T2w)	0.447±0.011	0.434±0.022

Table 5. The quantitative results based on the CBD datasets. Bold font denotes that the method performed the best among the considered methods.

4.4. Performance when using the multicenter traveling adult subjects with inverse-PE dataset

Finally, we evaluated the performance of the SACNet models with a public multicenter dataset that contains MR images from three traveling subjects collected over 10 sites. We first assessed whether our model obtained excellent SA correction quality after fine-tuning using only a small dataset (brain images of all subjects from the first site were used for model fine-tuning based on the models trained with the HCP dataset; see detailed fine-tuning strategy in SI-6). The quantitative comparisons presented in Fig. 6A and Table 6 show that the SACNet(T1w) approach significantly outperformed the Topup and S-Net methods in terms of FA-MSD (all $t \leq -10.433$ and all $p<0.001$, paired t test) and FA-SS (all $t \geq 3.149$ and all $p \leq 0.004$, paired t test). Detailed statistical results are provided in Supplementary Table S5.

To test whether our model could reduce multicenter effects, we calculated the mean coefficient of variation (CV) and regional intraclass correlation coefficient (ICC) of all corrected FA images for each SAC model separately (detailed definitions of CV and ICC are presented in SI-5). Lower CVs or higher ICCs represent smaller multicenter residuals. The SACNet(T1w) model obtained a significantly lower CV in all subjects (all $t \leq -12.674$ and all $p<0.001$ for all subjects, paired t test) than Topup and S-Net. Moreover, the SACNet(T1w) model obtained significantly higher ICCs in the frontal, temporal, and occipital lobes (all $t \geq 18.059$ and all $p<0.001$, paired t test) than Topup

and S-Net. Notably, we found that SACNet(T1w) obtained significantly higher ICC values than SACNet(wos) in the frontal (t=12.789, p<0.001, paired t test), temporal (t=6.806, p<0.001, paired t test), and subcortical (t=2.065, p=0.039, paired t test) areas, which indicated that the introduction of prior neuroanatomical information in SAs correction is valuable for reducing the multicenter effects. Detailed statistical results are provided in Supplementary Tables S6 and S7.

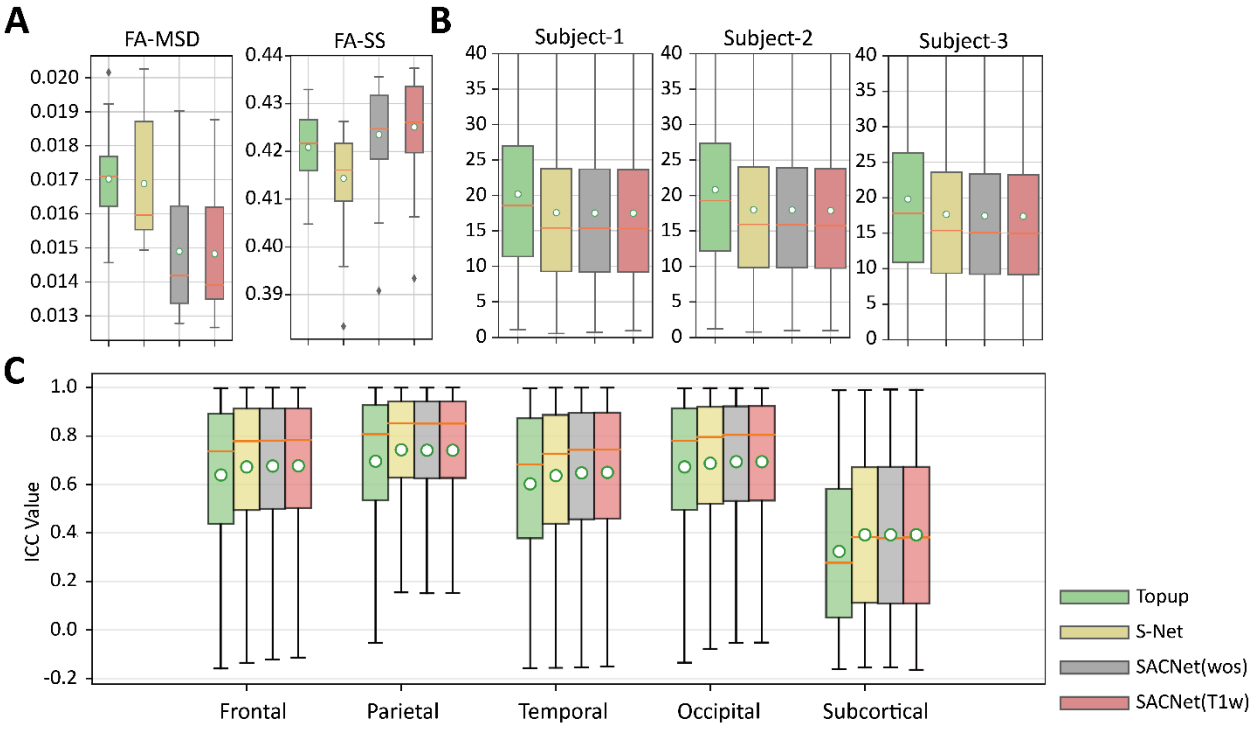


Fig. 6. (A) Boxplots of the FA-MSD and FA-SS values obtained by each SAC method in the multicenter dataset. (B) Boxplot of the CV distribution for each method in three traveling subjects. (C) Boxplot of the ICC distribution in each brain lobe for each method. For all the boxplots in (A), (B) and (C), the white circle indicates the mean value, and the coral horizontal line indicates the median value.

Method	Metrics		
	FA-MSD (1e-2)	FA-SS	NFV
Topup	1.703±0.123	0.421±0.008	9.80±9.23
S-Net	1.689±0.172	0.414±0.009	35.30±40.66
SACNet(wos)	1.490±0.185	0.423±0.010	**0.00±0.00**
SACNet(T1w)	**1.483±0.175**	**0.425±0.010**	0.27±0.57

Table 6. The quantitative results based on the multicenter dataset. Bold font denotes that the method performed the best among the considered methods.

4.5. Ablation studies

We separately assessed the effectiveness of the DEW module, DPF, CTF training and inference protocols (Section 4.5.1) and each component loss in the prior neuroanatomical information loss (Section 4.5.2) with ablation studies. The experiments were carried out based on the HCP dataset, and the T2w images were employed as the prior neuroanatomical information.

4.5.1. Ablation studies on the DEW module, DPF and CTF protocols

The effects of different combinations of the three components are presented in Table 7. Without the DEW module, the proposed method achieved the worst SAC performance, with the highest FA-MSD and lowest FA-SS values. Without the DPF, the number of folding voxels increased, and SAC performance decreased. Without the CTF protocols, SAC performance decreased, with increased FA-MSD and decreased FA-SS values. Additionally, SACNet required approximately 24 hours for training without the CTF protocols and only required approximately 12 hours for training with the CTF protocols.

4.5.2. Ablation studies on each component loss in $\mathcal{L}_{struct}$

$\mathcal{L}_{struct}$ contains the overall shape structural similarity loss $\mathcal{L}_{str-overall}$ and the pairwise structural similarity loss $\mathcal{L}_{str-pair}$. Table 8 shows that the SAC performance of the model depended on each component loss in $\mathcal{L}_{str}$. The best performance was achieved by leveraging both $\mathcal{L}_{str-overall}$ and $\mathcal{L}_{str-pair}$ in the network optimization. Furthermore, $\mathcal{L}_{str-overall}$ and $\mathcal{L}_{str-pair}$ both improved the SAC performance independently.

DEW	DPF	CTF	Metrics		
			FA-MSD (1e-2)	FA-SS	NFV
✗	✓	✓	2.541±0.335	0.297±0.025	70.80±77.782
✓	✗	✓	2.000±0.428	0.320±0.031	673.85±595.99
✓	✓	✗	2.077±0.429	0.321±0.029	18.70±33.28
✓	✓	✓	**1.978±0.420**	**0.322±0.030**	**15.22±38.17**

Table 7. The quantitative results of the ablation study based on the use of the DEW module, DPF and CTF protocols.

$\mathcal{L}_{str-overall}$	$\mathcal{L}_{str-pair}$	Metrics		
		FA-MSD (1e-2)	FA-SS	NFV
✗	✗	2.244±0.444	0.313±0.031	**13.62±27.60**
✗	✓	2.073±0.439	0.318±0.031	13.90±37.33

✓	✗	2.032±0.417	0.320±0.030	20.60±42.06
✓	✓	**1.978±0.420**	**0.322±0.030**	15.22±38.17

Table 8. The quantitative results of the ablation study of $\mathcal{L}_{str}$.

4.6. Runtime analysis

Table 9 shows the running time for estimating the inhomogeneity field image of a single scan based on each dataset for three SA correction methods: SACNet and two conventional methods, Fieldmap and Topup. The results show that SACNet is significantly more efficient than the conventional methods. These results indicate that SACNet has a significant advantage in processing large-scale datasets due to its ultrafast computational speed compared to those of conventional methods.

Method	HCP	HCP-D	dHCP	CBD	Multicenter
Fieldmap	—	—	—	~4 hours	—
Topup	3061	1740	986	—	1320
SACNet (CPU)	3.87	3.57	1.30	0.89	3.22
SACNet (GPU)	2.17	1.52	0.73	0.46	1.30

Table 9. Running time (in seconds). "—" denotes that there is no valid value.

5. Discussion and conclusion

We proposed an unsupervised multiscale convolutional registration network (SACNet) to remove SAs in brain EPI images. This model could generate diffeomorphic inhomogeneity fields based on either inverse-PE or single-PE images and employ prior neuroanatomical constraints from additional T1w or T2w images. Extensive experiments on neonatal, child and adult brain dMR images with different PE directions and PE numbers showed that our SACNet not only outperformed most popular conventional correction methods, such as Topup and Fieldmap, but also surpassed deep-learning based methods, such as S-Net and VoxelMorph. Furthermore, by a fine-tuning strategy with few samples in a multicenter dataset with traveling subjects, our model showed both better SAC performance and lower multicenter effects than the Topup and S-Net approaches. Our model reduced the time to generate inhomogeneous field images from the tens of minutes needed for conventional iterative approaches to a few seconds while maintaining state-of-the-art SAC performance; this outcome shows the potential advantages of SACNet for integrating multisite neuroimaging data in future brain development studies.

5.1. Applications to large-scale neuroimaging studies

Recent brain neuroimaging investigations have entered the era of "big data" (Bethlehem et al., 2022; Landhuis, 2017; Rutherford et al., 2022; Sejnowski et al., 2014; Xia and He, 2017) by integrating tens of thousands of image scans acquired at multiple centers. We emphasize that our SACNet approach is well suited for large-scale neuroimaging studies involving many individual scans for several reasons. First, we provided a range of models that have been pretrained based on diverse datasets with different ages and acquisition protocols, enabling users to fine-tune the models with only a few images according to their needs and to achieve excellent SAC in their own datasets (Section 4.4). Although recent cohort projects have used uniform EPI phase encoding protocols, many legacy datasets were acquired with various EPI protocols or even no EPI artifact correction sequences. Considering the high cost of acquiring human brain MR images, the utilization of existing databases is highly valuable. Second, our SACNet model effectively reduced the potential multicenter effects related to SAs (Section 4.3). Recent approaches for multicenter effect correction in brain MR images have received much methodological attention. Our model significantly reduced multicenter noise without using additional correction algorithms, highlighting the necessity of considering SAs in multicenter correction frameworks. Notably, our approach does not require correcting the multicenter effects in raw structural images, and further incorporation with multicenter structural image harmonization algorithms (Tian et al., 2022) could be attempted. Third, compared with conventional iterative optimization methods, SACNet can process many images from multiple subjects at fast speeds due to its ultrafast inference time (Section 4.6). Notably, SACNet does not require a large amount of CPU memory (approximately 3000 MB), making it convenient for batch processing in computing clusters. Fourth, we developed a comprehensive dMRI preprocessing pipeline specifically for SACNet, which integrated the output interface of SACNet with the input interface of existing dMRI postprocessing pipelines (Glasser et al., 2013). Finally, compared with existing deep-learning based SAC methods, for which only the developmental source code is available, we utilized the containerization technique to integrate our software source code and development environment, making it easy for users to deploy SACNet in their computational servers.

5.2. Effective network designs and integrated loss constraints in SACNet enable excellent SAC performance

A common deep-learning based brain registration framework is insufficient for solving the SAC

problem. Thus, we proposed several key designs to ensure the high quality of the generated inhomogeneity field and carefully evaluated their effectiveness. First, the combination of the DEW module and the DPF jointly enhanced the SAC capability of SACNet (first row vs. last row, second row vs. last row in Table 7). The DEW module successfully removed SAs by multiplying the Jacobian determinant of the inhomogeneity field with the geometric-corrected image, which facilitated the convergence of the network. In addition, the DPF constrained the inhomogeneity field in diffeomorphic space, thus reducing the number of invalid voxels (presented as negative intensity values and folding patterns) included in the DEW calculation process during model training and preventing overfitting of the network. Second, due to the severe SAs at temporal and frontal cortical boundaries (especially at the temporal pole and orbitofrontal cortex), anatomical morphologies within certain brain locations were barely conserved. Thus, it is imperative to incorporate prior neuroanatomical information to obtain excellent correction results. Previous methods that used structural images as additional inputs to the network were not sufficient to obtain good morphological images since these methods only provide information features and do not contribute to the loss function calculation (Hu et al., 2020; Schilling et al., 2020). To address this issue, we carefully designed a prior neuroanatomical information loss function, $\mathcal{L}_{struct}$, that was optimized for SAC by incorporating gradient-based information from structural images. This function $\mathcal{L}_{struct}$ includes two components: $\mathcal{L}_{str-pair}$ and $\mathcal{L}_{str-overall}$. $\mathcal{L}_{str-pair}$ was used to align the corrected image pair E_{PE_1} and E_{PE_2} to the structural image I_{struct} in a pairwise manner (third row vs. last row in Table 8), and $\mathcal{L}_{str-overall}$ improved the overall structural alignment between the final corrected image E_{final} and I_{struct} (second row vs. last row in Table 8). Moreover, the choice of an intensity-irrelevant structural metric, the NGF, allows users to use either T1w or T2w images as input neuroanatomical information, thereby improving the compatibility of SACNet for different types of clinical datasets. Third, the well-designed CTF training and inference protocols adopted in SACNet accelerated the training process and improved model convergence (third row vs. last row in Table 7), and similar strategies have been broadly deployed in conventional SAC methods (Bhushan et al., 2015; Duong et al., 2020a; Irfanoglu et al., 2015; Ruthotto et al., 2012).

5.3. Comparison with deep-learning based methods

Previous studies have proposed several deep-learning based registration approaches to address the SAC problem. For example, Bian et al. proposed correcting distortions by registering distorted b0

images in a single-PE direction to T1w images through the VoxelMorph backbone by optimizing the mutual information (MI) loss (Bian et al., 2023). However, this approach is limited to single-PE type data and was not compared with other methods designed specifically for the SAC problem. Duong et al. and Zahneisen et al. predicted inhomogeneity fields to remove SAs with 3D and 2D CNNs, respectively (Duong et al., 2020b; Zahneisen et al., 2020). However, the performance of these models is limited, as the models either ignore the intensity distortion problem or treat each volume slice as an independent example for training, resulting in inadequate SAs correction or inconsistent alignment between slices. Other methodological approaches for solving the SAC problem have also been developed. Several studies have used image generation approaches for SAC tasks. For example, Hu et al. and Ye et al. used high-resolution distortion-free point spread function encoded EPI (PSF-EPI) data as undistorted ground truth data for CNN training (Hu et al., 2020; Ye et al., 2023). Schilling et al. synthesized undistorted b0 images with U-Net or generative adversarial networks (GANs) and then entered both the "synthesized" and "real" b0 images as input into Topup to remove SAs (Schilling et al., 2020; Schilling et al., 2019). These supervised approaches need ground truth images as learning targets, which largely depend on the feature distribution of the training images. This may lead to difficulty when facing brain images with heterogeneous appearance, such as those of neonatal brain scans. Moreover, when these approaches are applied to a new neuroimaging dataset, acquiring undistorted images for fine-tuning can often be a costly endeavor. In contrast to image generation-based methods, unsupervised registration-based methods, such as our approach, require no ground truth labels. Interestingly, Qiao et al. proposed the distortion correction network (DrCNet) by feeding fiber orientation distribution (FOD) information into U-Net and successfully corrected residual distortions that could not be eliminated by Topup (Qiao and Shi, 2021). Compared to DrCNet, our SACNet method can be applied not only to dMRI data but also to fMR images. Moreover, the performance of SACNet may be further enhanced by incorporating rich diffusion-based information, such as DWIs and FODs, into the integrated loss function.

5.4. Limitations and future directions

Several issues in this study should be considered. First, although the additional neuroanatomy priors improve the SAC performance of our SACNet, it may be worthwhile to explore the further employment of white matter information from dMRI data itself (Irfanoglu et al., 2015; Qiao and

Shi, 2021; Qiao et al., 2019). Second, the presence of other nonnegligible artifacts, such as eddy current-induced distortions and intrasubject movements, in dMRI data should be acknowledged (Andersson and Sotiropoulos, 2016). It would be interesting to develop a deep-learning based tool in conjunction with SACNet to address these artifacts. Third, registration-based methods may become unsatisfactory in 7T MR images due to severe signal loss issues. Given that the deep generative model (DGM) has shown the ability to capture complex distributions of real 7T data (Nie et al., 2018), combining the DGM with SACNet could be a promising approach. We hope that SACNet can offer a general framework for SAC task in multicenter datasets with top-ranking performance, robust output and efficient computational speed, which could facilitate a wide variety of future brain studies using large-scale multicenter neuroimaging datasets.

Data availability:

The datasets from the Human Connectome Project and the Lifespan Human Connectome Project Development are available at https://www.humanconnectome.org. The dataset from the Developing Human Connectome Project is available at https://www.developingconnectome.org. The Multicenter dataset is available at https://doi.org/10.6084/m9.figshare.8851955.v6. Raw imaging data is available from the corresponding authors upon reasonable request.

Code availability:

The source code that implements our software is available at https://github.com/RicardoZiTseng/SACNet.

Acknowledgments:

The study was supported by the National Natural Science Foundation of China (Nos. 31830034, 82021004, 81801783, 82102139 and 82202243) and the Changjiang Scholar Professorship Award (T2015027). We are grateful to the Developing Human Connectome Project (DHCP), the Human Connectome Project (HCP) and the Lifespan Human Connectome Project Development (HCPD). HCP imaging data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research and by the Mc-Donnell Center for Systems Neuroscience at Washington University. The Lifespan HCP imaging data were supported by the National Institute of Mental Health of the National Institutes of Health under Award Number U01MH109589 and by funds provided by the McDonnell Center for Systems Neuroscience at Washington University in St. Louis. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions:

Z.L.Z., T.D.Z., and Y.H. designed the research; W.W.M., Y.P.W., R.C., H.B.Z., S.P.T., J.H.G., S.Z.Q., Q.Q.T., H.J.H., S.T., Q.D., and Y.H. collected the imaging dataset; J.Y.Z., X.Y.L., L.L.S., Y.H.Z., and Y.H. provided the methodological instruction; Z.L.Z., and T.D.Z. performed the data analysis; Z.L.Z., and T.D.Z. wrote the paper; Z.L.Z., T.D.Z., and Y.H. revised the paper.

Conflicting Interests: The authors have declared that no conflicting interests exist.

References

Andersson, J.L., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20, 870-888.

Andersson, J.L., Sotiropoulos, S.N., 2016. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. Neuroimage 125, 1063-1078.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. Voxelmorph: a learning framework for deformable medical image registration. IEEE transactions on medical imaging 38, 1788-1800.

Bethlehem, R.A., Seidlitz, J., White, S.R., Vogel, J.W., Anderson, K.M., Adamson, C., Adler, S., Alexopoulos, G.S., Anagnostou, E., Areces-Gonzalez, A., 2022. Brain charts for the human lifespan. Nature 604, 525-533.

Bhushan, C., Haldar, J.P., Choi, S., Joshi, A.A., Shattuck, D.W., Leahy, R.M., 2015. Co-registration and distortion correction of diffusion and anatomical images based on inverse contrast normalization. Neuroimage 115, 269-280.

Bian, Z., Shao, M., Carass, A., Prince, J.L., 2023. DrDisco: Deep Registration for Distortion Correction of diffusion MRI with single phase-encoding, Medical Imaging 2023: Image Processing. SPIE, pp. 292-296.

Biswal, B., Zerrin Yetkin, F., Haughton, V.M., Hyde, J.S., 1995. Functional connectivity in the motor cortex of resting human brain using echo‐planar MRI. Magnetic resonance in medicine 34, 537-541.

Bowtell, R., McIntyre, D., Commandre, M., Glover, P., Mansfield, P., 1994. Correction of geometric distortion in echo planar images, Soc. Magn. Res. Abstr, p. 411.

Chang, H., Fitzpatrick, J.M., 1992. A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. IEEE Transactions on medical imaging 11, 319-329.

Duong, S., Phung, S.L., Bouzerdoum, A., Taylor, H.B., Puckett, A., Schira, M.M., 2020a. Susceptibility artifact correction for sub-millimeter fMRI using inverse phase encoding registration and T1 weighted regularization. Journal of Neuroscience Methods 336, 108625.

Duong, S.T., Phung, S.L., Bouzerdoum, A., Schira, M.M., 2020b. An unsupervised deep learning technique for susceptibility artifact correction in reversed phase-encoding EPI images. Magnetic Resonance Imaging 71, 1-10.

Erkol, H., Demiralp, E., 2007. The Woods–Saxon potential with point interactions. Physics Letters A 365, 55-63.

Fu, Y., Lei, Y., Wang, T., Curran, W.J., Liu, T., Yang, X., 2020. Deep learning in medical image registration: a review. Physics in Medicine & Biology 65, 20TR01.

Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., 2013. The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage 80, 105-124.

Haber, E., Modersitzki, J., 2007. Intensity gradient based registration and fusion of multi-modal images. Methods of information in medicine 46, 292-299.

Hagmann, P., 2005. From diffusion MRI to brain connectomics. EPFL.

Hédouin, R., Commowick, O., Bannier, E., Scherrer, B., Taquet, M., Warfield, S.K., Barillot, C., 2017. Block-matching distortion correction of echo-planar images with opposite phase encoding directions. IEEE Transactions on Medical Imaging 36, 1106-1115.

Holland, D., Kuperman, J.M., Dale, A.M., 2010. Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging. Neuroimage 50, 175-183.
Hu, Z., Wang, Y., Zhang, Z., Zhang, J., Zhang, H., Guo, C., Sun, Y., Guo, H., 2020. Distortion correction of single-shot EPI enabled by deep-learning. Neuroimage 221, 117170.
Irfanoglu, M.O., Modi, P., Nayak, A., Hutchinson, E.B., Sarlls, J., Pierpaoli, C., 2015. DR-BUDDI (Diffeomorphic Registration for Blip-Up blip-Down Diffusion Imaging) method for correcting echo planar imaging distortions. Neuroimage 106, 284-299.
Jaderberg, M., Simonyan, K., Zisserman, A., 2015. Spatial transformer networks. Advances in neural information processing systems 28.
Jezzard, P., Balaban, R.S., 1995. Correction for geometric distortion in echo planar images from B0 field variations. Magnetic resonance in medicine 34, 65-73.
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Landhuis, E., 2017. Neuroscience: Big brain, big data. Nature 541, 559-561.
Lerch, J.P., Van Der Kouwe, A.J., Raznahan, A., Paus, T., Johansen-Berg, H., Miller, K.L., Smith, S.M., Fischl, B., Sotiropoulos, S.N., 2017. Studying neuroanatomy using MRI. Nature neuroscience 20, 314-326.
Littlejohns, T.J., Holliday, J., Gibson, L.M., Garratt, S., Oesingmann, N., Alfaro-Almagro, F., Bell, J.D., Boultwood, C., Collins, R., Conroy, M.C., 2020. The UK Biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions. Nature communications 11, 2624.
Makropoulos, A., Robinson, E.C., Schuh, A., Wright, R., Fitzgibbon, S., Bozek, J., Counsell, S.J., Steinweg, J., Vecchiato, K., Passerat-Palmbach, J., 2018. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. Neuroimage 173, 88-112.
Nie, D., Trullo, R., Lian, J., Wang, L., Petitjean, C., Ruan, S., Wang, Q., Shen, D., 2018. Medical image synthesis with deep convolutional adversarial networks. IEEE Transactions on Biomedical Engineering 65, 2720-2730.
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., 2019. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems 32.
Qiao, Y., Shi, Y., 2021. Unsupervised deep learning for FOD-based susceptibility distortion correction in diffusion MRI. IEEE Transactions on Medical Imaging 41, 1165-1175.
Qiao, Y., Sun, W., Shi, Y., 2019. FOD-based registration for susceptibility distortion correction in brainstem connectome imaging. NeuroImage 202, 116164.
Reber, P.J., Wong, E.C., Buxton, R.B., Frank, L.R., 1998. Correction of off resonance - related distortion in echo - planar imaging using EPI - based field maps. Magnetic Resonance in Medicine 39, 328-330.
Rutherford, S., Fraza, C., Dinga, R., Kia, S.M., Wolfers, T., Zabihi, M., Berthet, P., Worker, A., Verdi, S., Andrews, D., 2022. Charting brain growth and aging at high spatial precision. elife 11, e72904.
Ruthotto, L., Kugel, H., Olesch, J., Fischer, B., Modersitzki, J., Burger, M., Wolters, C., 2012. Diffeomorphic susceptibility artifact correction of diffusion-weighted magnetic resonance images. Physics in Medicine & Biology 57, 5715.

Schilling, K.G., Blaber, J., Hansen, C., Cai, L., Rogers, B., Anderson, A.W., Smith, S., Kanakaraj, P., Rex, T., Resnick, S.M., 2020. Distortion correction of diffusion weighted MRI without reverse phase-encoding scans or field-maps. PLoS One 15, e0236418.

Schilling, K.G., Blaber, J., Huo, Y., Newton, A., Hansen, C., Nath, V., Shafer, A.T., Williams, O., Resnick, S.M., Rogers, B., 2019. Synthesized b0 for diffusion distortion correction (Synb0-DisCo). Magnetic resonance imaging 64, 62-70.

Sejnowski, T.J., Churchland, P.S., Movshon, J.A., 2014. Putting big data to good use in neuroscience. Nature neuroscience 17, 1440-1441.

Somerville, L.H., Bookheimer, S.Y., Buckner, R.L., Burgess, G.C., Curtiss, S.W., Mirella, D., Stine, E.J., Gaffrey, M.S., Harms, M.P., Cynthia, H., 2018. The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds. NeuroImage 183, S1053811918307481-.

Sporns, O., Tononi, G., Kötter, R., 2005. The human connectome: a structural description of the human brain. PLoS computational biology 1, e42.

Studholme, C., Constable, R.T., Duncan, J.S., 2000. Accurate alignment of functional EPI data to anatomical MRI using a physics-based distortion model. IEEE transactions on medical imaging 19, 1115-1127.

Tax, C.M., Bastiani, M., Veraart, J., Garyfallidis, E., Irfanoglu, M.O., 2022. What's new and what's next in diffusion MRI preprocessing. NeuroImage 249, 118830.

Tian, D., Zeng, Z., Sun, X., Tong, Q., Li, H., He, H., Gao, J.-H., He, Y., Xia, M., 2022. A deep learning-based multisite neuroimage harmonization framework established with a traveling-subject dataset. NeuroImage 257, 119297.

Tong, Q., He, H., Gong, T., Li, C., Liang, P., Qian, T., Sun, Y., Ding, Q., Li, K., Zhong, J., 2020. Multicenter dataset of multi-shell diffusion MRI in healthy traveling adults with identical settings. Scientific Data 7, 157.

Turner, R., Le Bihan, D., Maier, J., Vavrek, R., Hedges, L.K., Pekar, J., 1990. Echo-planar imaging of intravoxel incoherent motion. Radiology 177, 407-414.

Warach, S., Gaa, J., Siewert, B., Wielopolski, P., Edelman, R.R., 1995. Acute human stroke studied by whole brain echo planar diffusion‐weighted magnetic resonance imaging. Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society 37, 231-241.

Xia, M., He, Y., 2017. Functional connectomics from a "big data" perspective. Neuroimage 160, 152-167.

Yamashita, A., Yahata, N., Itahashi, T., Lisi, G., Yamada, T., Ichikawa, N., Takamura, M., Yoshihara, Y., Kunimatsu, A., Okada, N., 2019. Harmonization of resting-state functional MRI data across multiple imaging sites via the separation of site differences into sampling bias and measurement bias. PLoS biology 17, e3000042.

Ye, X., Wang, P., Li, S., Zhang, J., Lian, Y., Zhang, Y., Lu, J., Guo, H., 2023. Simultaneous superresolution reconstruction and distortion correction for single‐shot EPI DWI using deep learning. Magnetic Resonance in Medicine.

Zahneisen, B., Baeumler, K., Zaharchuk, G., Fleischmann, D., Zeineh, M., 2020. Deep flow-net for EPI distortion estimation. Neuroimage 217, 116886.